analysis of statistical datasets and actors of the quantification of tourism. methodological propositions

projet «overtourism? les villes comptent»
working paper n°1

Valérian Geffroy, Victor Anduze Rivero et Davide Ceccato
avec la collaboration de Fabrice Bardet, Roser Cussó, Giovanni Favero et Mathis Stock

avril 2022

anr agence nationale de la recherche
Fonds national suisse
ENTPE
UNIL | Université de Lausanne

Analysis of statistical datasets and actors of the quantification of tourism—methodological propositions

Valérian Geffroy, Victor Anduze Rivero et Davide Ceccato
Avec la collaboration de Fabrice Bardet, Roser Cussó, Giovanni Favero et Mathis Stock

Table des matières

Introduction 2
1. Grounding methods in theories and disciplines 4
1.1. Sociology of quantification 5
1.2. Geography 7
1.3. Digital studies 9
2. "Mapping" actors and datasets of the quantification of tourism 13
2.1. Exploration: interviews and documentary collection 13
2.2. Inventory and mapping 14
2.3. Use of the mapping after the exploratory stage 18
3. Questioning actors, datasets, and discourses 18
3.1. Semi-structured interviews 19
3.2. Ethnomethodology and ethnostatistics 21
3.3. Participant observation in urban tourism space 23
3.4. Analysing discourses on statistics 24
Conclusion and pending questions 24
References 25

Introduction

How exactly is tourism quantified? There are many levels of answer to this question, and our research project aims at the most detailed of these levels. It does not aim primarily at an exhaustive description of every single dataset, actor or process involved, although we will need to give a good sense of the main elements of local, regional, or national "tourism quantification systems" in our case studies. But it aims mostly at a critical interrogation of those datasets, actors, and processes, in terms of their social and political embeddings. It is an answer to "how" tourism is quantified in that it unfolds the practices and decisions of human actors who make or use the statistics of tourism.

But for any critical study of quantification in social sciences, the question "how?" is not enough, or not precise enough. It contains all the following questions:

- *Who* makes the statistics? Organisations, companies, people: what are their aims, their jobs, their skills, their training?
- *What* are these statistics? What are they supposed to count? In what forms (databases, surveys, administrative registries, digital footprints…)?
- *When* are they made? When were the indicators and methods designed? With what frequency are the data collected and made available? On what period of time are they meant to bring knowledge?
- *Where* are they made? In which national, cultural, or political context? At what scale, on what areas or locations?
- *How* are they made? With what conventions, what mathematical calculations and technical processes, by what professional groups, what tools, in what times and spaces… How are they published, if published at all? In a 'raw', processed, or aggregated form? Through what channels, media outlets, intermediaries?
- *Why* are these statistics produced? This includes on the one hand the question of the origins and principles of statistics production, and on the other hand their ends and uses, be they technical, political, or even polemical.

In addition, our project raises two questions that are directly related to two contemporary issues which interrogate tourism statistics with a renewed acuteness:

- How, in the context of the overtourism controversy, is the statistical argument used by cities and their citizens when addressing tourism?
- Does the rise of digital and/or big data, and of the producers of these data, create new forms of governance of cities and tourism?

This document proposes a methodological protocol to answer these questions on the specific topic of the quantification on tourism. Figure 1 illustrates how this protocol is integrated to our general research design, in particular how it will be applied to every case study and geographical level of study.

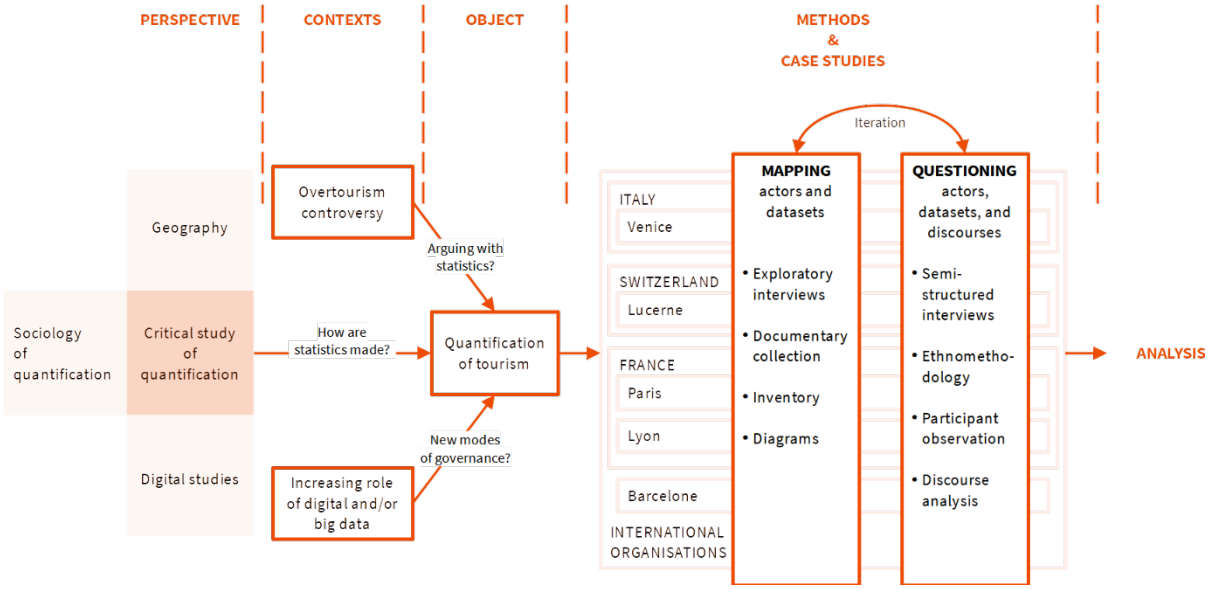

Figure 1: Research design

This project is inspired by several previous works on the quantification of tourism (Pratt & Tolkach, 2018; Stock et al., 2017, chapter II; Terrier, 2006), but it aims to be more encompassing, more developed, and more up-to-date. Indeed, these previous works were primarily focused on the most authoritative statistics (UNWTO, national statistical institutes), and their critiques were mostly articulated around the matters of reliability, incompleteness, and adequacy to the needs of research, thus addressing in general the finer technical problems. In contrast, we propose an entire research protocol dedicated to the critical study of the quantification of tourism, understood as a broad field of actors and datasets, and where the technicalities of statistics are recontextualised within organisations, politics, debates, work practices and processes, social uses, etc. This research protocol is more encompassing in that, for instance, it does not only question the statistical variables (hotel nights, international visitor arrivals…) but also the broader dimensions of tourism that they quantify (tourists' presence, tourists' impacts…). This research protocol is more systematic in that it includes a wide range of geographical scales, from the local or micro-local to the global; it is also more systematic in that it attempts to 'map' entire quantification systems at these different scales, by drawing an inventory of all key actors and datasets. This requires keeping a broad and open perspective about the definition of tourism and tourists, at least in the first phases of the project. Indeed, the study of concurrent or conflicting definitions of tourism based on different statistics, or the potential failure of current definitions to grasp the current shifts in tourism practices, will be important objects of study for this project. Finally, this project directly grapples with the latest issues of tourism to understand how they influence—by fueling them, or unsettling them, for instance—the systems of quantification. In particular, we include private companies, most importantly ICT companies (digital platform, telecoms), as new key players in the quantification systems; and we include anti-tourism movements and preoccupations in the political balances that may affect the quantification of tourism.

The multiscalar perspective of our research also leads us to analyse a stratum of actors in the quantification of tourism that has been little explored in the previously mentioned works. By looking at the local scale (the city/metropolis) we find ourselves at a scale where the production of statistics on tourism does not have a pre-established framework, as is the case at the international level where the "equivalence conventions" are defined by the OECD, UNWTO or EUROSTAT. At the national level, it is often the central institutions that establish the statistical standards. The role of cities is therefore not well defined if we analyse the strict "production" of data, since mixed or *ad hoc* standards can be applied. However, the existing bibliography on the

subject of tourism quantification has often privileged the analysis of methods implemented by international and national institutions, since they are identifiable thanks to the manuals that these institutions regularly publish. The critical analysis of the circulation of data will allow us to take into account the local scale, which would otherwise be ignored if we were content to look only at the statistical apparatus on which a fairly large bibliography already exists.

In a first section of this document, we lay out the main theories and disciplines that inform our methodological propositions (see "perspective" on Figure 1). In the next sections, we present two main groups of methods. The first group—"mapping" actors and datasets—mostly answers the 'descriptive' questions, who, what, when, and where (section 2). The second group of methods—"questioning" actors, datasets, and discourses—furthers and deepens these answers as well as it addresses the "analytical" questions: how and why (section 3).

1. Grounding methods in theories and disciplines

Our research is focused on the study of statistical data production, including the function of statistics as a tool for evidence (objectification) and coordination (decision) (Desrosières, 2014). But it is also deployed on different case studies, at different spatial scales, and anchored in the contemporary debates on tourism in the city. Hence, the research design requires a variegated methodological toolset, adapted to our main theoretical perspectives and disciplinary interests, but also to the specific field and issues we are dealing with. Thus, we first outline the main approaches that combine to define our specific perspective, and show how these approaches help build an adequate and complete methodological toolset. This toolset should allow to explore, and then describe and explain, the role of the different actors and techniques that make up the structure (in the broadest sense) of tourism quantification; it should also allow to understand how practices and discourses of quantification are involved in debates and controversies, chiefly on the issue of overtourism.

Three main approaches are presented below: sociology of quantification, geography, and digital studies. Together, they make a large part of the theoretical framework of the research project. However, the goal in this document is not to fully lay out this theoretical framework, but only to extract from it the tools, themes, or perspectives, that are useful—or necessary—to construct an adequate methodological protocol. Thus, section 1 develops as follows. The sociology of quantification (1.1.) offers a critical stance to the social nature of statistic-making, through an attention to techniques of quantification, interactions between actors, and to professional statistical praxis. It also points to the necessity of studying the circulation of statistical discourses, which will be of particular importance given our interest in the controversy on overtourism. Geography (1.2.) highlights the spatial stakes and issues of the quantification of social phenomena, and of mobilities in particular; it thus emphasises the need to make our methods sensitive to questions of scale, delimitation, and location. Besides, extensive work has been conducted in geography on tourism in the city and the related issues, giving precious insights into the matters of sharing space, of differentiated relations to space, and power relations between the various inhabitants. Digital studies (1.3.) are to be taken into account in any study of contemporary statistics, because many new sources of quantification are now natively digital, as the "footprints" of activity on digital devices have opened a whole new world of data. And on the critical side, digital studies have coined many concepts and methods to examine the effects these technological evolutions have on social, economic, and political balances—in our case, regarding the professional field of quantification, tourism practices, and urban governance.

1.1. Sociology of quantification[1]

The use of statistical data (from registers and surveys, and more recently from so-called big data) to describe society has been the subject of many debates and controversies throughout history. There have been many arenas of discussion about the method for quantifying various features of a society, about the definition and delimitation of these features and the metrology for expressing their magnitudes, but they can be grouped into two large groups. The first is that which corresponds to the sphere of action/power. For the first administrative statistical services, formalized in the 18th century in countries such as France, Prussia and Great Britain (Desrosières, 2010: 180), the method for counting the population, estimating agricultural production or calculating the nation's wealth was not self-evident, nor was the way to represent their respective quantities. The figures made available to the government, to rationally orientate decision making, have therefore been the object of multiple disagreements. Methods of quantification have appeared, disappeared and reappeared according to the historical conjecture of each State at different moments of its history. This is still a very current phenomenon, whether we are talking about the ethnic division of a population (Bardet, 2012; Jugnot, 2013), international statistics on education (Cussó, 2003), indicators for evaluating the cost-effectiveness of public policies (Desrosières, 2014), or, of course, tourism (Stock et al, 2017, Terrier, 2006).

These different approaches to the question of the most accurate method of quantification to account for the statement of the elements that make up a nation or a society have been the subject of lively debate in another *milieu*, that of statistical science. In his *Politique des grands nombres* (2010), Alain Desrosières illustrates numerous examples of these scientific disagreements that are strongly intertwined with political disagreements[2]. The study of these crucial moments in the evolution of statistics as a science, but also as evidence for decision making, is called the "historical sociology of quantification" (*Ibid.*). By delving into the sociological study of these past controversies, we can better understand their evolution and, sometimes, their present form. It also allows us to move beyond the controversy over the use of statistics as scientific evidence and as support for political decision-making. Indeed, it has served both purposes since its simultaneous appearance in the political and scientific spheres (etymologically it has always been the science of the state).

Rather than confronting the naturalistic character of statistics (it measures reality) or its constructivist character (it creates reality), sociology invites us to investigate the links between science (statistics) and power (which relies on numbers), and the tensions within each of these fields. To do so, we need to understand the "equivalence systems/conventions" that make commensurable the data resulting from quantification, and the technical and political means put in place to establish these equivalences. This opens up two perspectives for doing this sociology. One of them would be externalist, interested in the use that the different actors make of statistics, in the coordinating role of this tool. The other, of the internalist type, focuses on the contents "of knowledge itself, of instruments and results, of theorems and their demonstration [the proof role of the statistical tool]." (Desrosières, 2014: 253). However, the sociological understanding of quantification is not limited to one field or the other, but deals with the bridges woven between the two.

Who quantifies, through which methods, how were these methods designed, what resistance do they find, what reactions do the results obtained provoke, who participates in these

[1] In the Anglo-Saxon literature, Wendy Espeland and Mitchell Stevens (2008), propose the term "sociology of quantification" as an equivalent translation to the French *Sociologie de la quantification.* Previously, Espeland and Stevens (1998), proposed the term "commensuration studies". However, we retain here the first translation because of its linguistic and temporal proximity

[2] The replacement of De Ferrier (who advocated the use of ordered statics according to literary and descriptive rules) by Duvillard (who claimed the use of arithmetic tables as the main tool of quantification for administrative purposes) in 1806 is a good example, among others, of how methodological and technical differences were strongly linked to the administrative structure of the State) (Desrosières, 2010: 49).

processes? These are the questions that sociology invites us to ask ourselves when we work on or with data resulting from a process, always complex, of quantification.

1.1.1. Three work sites, three moments for analysis

In the work of the founder of the sociology of quantification in France, Alain Desrosières, but also in the work of Anglo-Saxon authors who do not claim to be doing a sociology of quantification, but who nevertheless place the production of data within the study of social and/or historical facts (Anderson, 2015; Hacking, 2008; Porter, 1985), we find three, so to speak, recurring work sites. The first is that of the delimitation of the object. In our case, it would be a matter of delimiting precisely what is meant by tourism, tourist, travel, tourist travel, etc.

Secondly, these works are concerned with observing the political conditions, the intellectual and social context (in a perspective of history of ideas and history of science) and the technical means existing at a given time, which allowed the appearance of an object to be quantified, the techniques for quantifying it and the allocation of the means to do so. In other words, they are interested in the genesis of an object and/or a process of quantification.

The third area is the selection of key controversies and, as with the study of genesis, the conditions in which they have developed. Which schools, which interest groups, which antagonistic or complementary ideas were the subject of, or participated in, these controversies? Often, these works also include a perspective on the current definition of the object and the contemporary, or even future, stakes of its quantification.[3]

These three projects are therefore built on three forms of analysis. The first, that is to say, the definition of the object, requires an epistemological perspective, because it is the work that serves as the foundation of what will be explained afterwards (in our case, it is the work that corresponds to the delimitation of the tourism/tourist object). Secondly, the work on the genesis of the object leads us to the study of its history but also to the analysis of the public policies that have allowed its appearance. And one could say the same for the study of controversies, as the important place of controversies in power relations forces us to sociologise their history. The sociology of quantification thus requires multiple competences, if we are guided by the works of the last 40 years that different authors (those cited here are non-exhaustive examples) have developed by mobilizing this perspective to better understand the stakes of the quantification of the world.

1.1.2. A cross-disciplinary approach

The sociology of quantification, from an empirical point of view, can be defined as a particular approach to the observation and study of social facts. This approach invites us to place numbers, but especially the technical, historical, social and political process that produces them, at the centre of sociological reflection. As we have seen, this work of objectifying the process of producing figures requires the mobilization of several disciplines. The sociology of quantification is therefore, in essence, an interdisciplinary approach. One of the proofs of this can be found in the diversity of profiles of the researchers who have contributed to this field of reflection: we find philosophers (Hacking, 2008), economists (Alonso & Starr, 1987) sociologists (Alonso & Starr, 1987; Espeland & Stevens, 2008), statisticians (Desrosières 2010 and 2014), historians (Porter, 1985), etc.

In what follows, the sociology of quantification should not be understood as one discipline among others that would allow us to obtain singular information on our research object. Rather, it will serve as a guideline for the critical analysis of statistical production related to tourism (by mobilizing its three "work sites" of research, explained above). Whether it is to understand the territorial scale at which the statistical production of tourism is carried out, the history of the

[3] The order of these work sites varies according to the authors. Here, they have been listed as such for explanatory convenience.

technical and political controversies of the quantification of tourist activity, or the economic and political stakes that shape it, the invitation made by the sociology of quantification to see the figures as a social fact (and thus to define them in a precise manner, to contextualize them, in short, to explain them) will remain a common trait in dealing with our research questions.

The production and dissemination of statistics related to tourism, as an object of study, invites us to think about the scale at which they are made (international, national, local), the development of techniques for doing so (statistics from censuses, sample surveys and big data) and the social reactions that mobilize them (social and political movements denouncing the excess of tourists). The sociology of quantification is therefore a very relevant approach to answer these questions, which intertwine various social actors and their relations to the production and dissemination of figures and statistics.

A "work field" that has not been much explored by the sociology of quantification is the daily practice of experts, technicians and managers in the production and dissemination of statistical figures. How do these people, through their level of competence, training, rank within the structures in which they work, social *milieu*, workplace, responsibilities and many other aspects linked to daily contact with the construction, production and dissemination of statistical data, conceive of the information with which they work on a daily basis? The question, as it has just been posed, leads us to a *bourdieusian* reflection on the *habitus* specific to the professions linked to statistical figures (these professions are very diverse, we are aware of this, but at this stage of the work we will remain broad in the range of possible professions to begin our investigations). Ethnostatistics (see section 3.2.) offers us an important set of tools to understand this "culture of working with numbers". When we go into the field to survey statisticians, managers, directors of institutions, ministers, employees of tourist offices, etc., the methodological perspective of ethnostatistics will undoubtedly be very relevant in order to make, in addition to a sociology of quantification, a sociology of quantifiers.

1.2. Geography

To a cross-disciplinary approach grounded in the sociology of quantification, geography contributes by shedding light on the spatial conditions and problems of quantification, especially the quantification of a mobile phenomenon such as tourism. And regarding tourism in cities, there is a wealth of geographical research on tourists' practices, movements, and diverse ways of relating to place; and on the issues and struggles that may arise when tourists and residents cohabit. Thus, geography is a key perspective for the critical analysis of the quantification of tourism, and for the understanding of how tourism affects the life and governance of cities.

1.2.1. The spatial issues of quantification

In all definitions of the concept, tourism involves moving from one place, defined as "home", to another place, qualified as "away", or simply "not home". This is an inherently spatial definition; and a definition that relies on several conventions of a spatial nature. For purposes of quantification especially, a place of residence must be attributed to every individual—this is a foundational convention for statistics in general, starting with census. Then, the types of movement that can enter the definition of tourism must be defined, for instance with a timeframe (longer than one day or including a night), or with a distance from home. Already problems arise for the quantification of such movements at the individual scale: geography, as a discipline equipped to analyse the various relations to space and practices involved in mobilities, has long emphasised the problematic nature of statistical definitions of tourism, which introduce rigid boundaries in the continuum formed by multiple forms of mobilities (Hall, 2005), subsume under a single category trips with motivations as different as business and vacation (Stock et al., 2017), and are ill-equipped to handle patterns of multi-residence (Duchêne-Lacroix et al., 2013).

The other main angle for the quantification of tourism is the counting of incoming flows in a given territory, or "destination". This is the main problem for our research object, tourism in cities. In all destinations, one of the central geographical questions in the quantification of tourism is the selection of the relevant area to consider the phenomenon (Stock et al., 2017). The measure of flows obviously leads to very different figures according to the different chosen surfaces and boundaries, and frequently leads even in official statistics to distorted or disputable views of tourism (Pratt & Tolkach, 2018; Terrier, 2006). Indeed, the figures are heavily dependent on the territorial "mesh" under scrutiny, especially on national borders, as they are the basis for the measure of international tourism. In general, the more administrative delimitations there are, the more important the counted flows are. Hence, the strict methodological nationalism of UNWTO's tourism statistics makes for a poor comparability between the dense puzzle of nations with open internal borders that is the EU, and continent-sized countries such as the USA or China. The potential biases and manipulations (for instance to favour a country or city in the highly symbolic rankings of tourism destinations) inherent to administrative delimitations are a first reason to pay close attention to reference territories of tourism statistics.

Another reason is the complexity of matching the measure of tourism with administrative units that constitute the basis of most statistical datasets. In particular, the spatial extension of the tourism activity in an urban area does not have a direct correspondence with the administrative boundaries of the city (Ostertag and Wöber, 2010: 30). Indeed, in terms of tourists' activity, the city does not exist as a delimited entity, but rather as a continuum with its neighbourhood (Stock, 2019: 6), or in relation with important nearby attractions (such as Versailles and Eurodisney for the Paris area). This frequently leads city governments to favour statistics at the scale of urban regions to maximise the figures of tourist frequentation. But efforts are being made on the European level to institute a standardised definition of cities and urban areas (Dijkstra & Poelman, 2014), which should lead to better comparability of urban tourism statistics. In the domain of the official definition of cities, geography works hand in hand with statistics. Concepts such as urban areas, city centres, suburbs, peri-urban areas, come from geography and are translated in statistical categories with indicators such as population or built density, and commuting trips. Geographers maintain a lively discussion with statisticians on these questions, which include a critique of how statistics support discourses and policies on urban and rural matters (Pistre & Richard, 2018).

1.2.2. Place, territory, and local arrangements of tourism

Analysing the spatial dimensions of tourism in cities helps putting back complexity in the phenomenon, which quantification inevitably simplifies. Moreover, understanding the complexity of the insertion of tourism in the urban fabric is imperative to grasp the local expressions of the overtourism controversy. Geographical discussions on place, territory, and more generally on the varied and changing relations of humans to space allow to introduce in the perspective the crucial questions of meaning, belonging, and power dynamics.

A first element of complexity that represents an important challenge to the quantification of tourism is the difficulty of estimating tourism activity and tourists at the local or micro-local scale, especially in short timeframes, and especially in large cities. These are by definition diverse and host at any time a wide range of people and activities, including many temporary visitors among which only a fraction may, in some cases, be unequivocally recorded as tourists by the use of dedicated services or infrastructures (short-term accommodation in particular). But for economic activity, or for regulation of flows or local impacts, the total amount of people present at a given time in a city, neigbourhood, or square, is precious information. Estimates can be calculated from general surveys, but only on broad territorial scales (Terrier, 2007). Mobile phone data represent a major novel avenue of work (see section 1.3.), with promises of much more accurate estimates of the presence of tourists in destinations (Cousin & Hillaireau, 2019; Vanhoof et al., 2017). Geographical research on tourism has seen many experiments to better track or estimate tourists' presences within cities, with qualitative surveys or monitoring devices (see for

instance Bauder et al., 2014; Shoval & Ahas, 2016). This is a particularly important avenue of research in that the micro-local distribution (at the level of street, block, or even finer) of tourist practices and movements is likely the most relevant to measure the concentration of the effects of tourism in cities: recent research in tourism geography has shown that most often, even in major tourist destinations, tourism activity and its potential disturbing effects is concentrated in rather small areas and specific hotspots (Camarillo-Naranjo et al., 2022; Eggli, 2021).

Furthermore, actors of the tourism sector and the quantification system have their own conceptions of space, their own relations with territorial delimitations and scales. Some of them (administrations for instance) have clearly defined geographical levels and ranges of action, while other can strategically shift or project their activity (transnational companies in particular). And the centrality of actors in the quantification systems relies in part on their ability to act on different territorial scales. National statistical institutes control the production of statistics on regional levels as well as the national level; transnational companies seek to reproduce their processes or adapt them at the local scale to provide a globally standardised offer—or tourist experience. Other actors play a key role on the local level, for instance as aggregators and compilers of relevant statistics for the tourism activity in a specific city. Keeping in mind the duality between the scale of action and scale and quantification also allows to be alert to the ways the actors may play with scales to serve their interests or their discourse, as is the case when Airbnb puts forward figures relative to a whole municipality or urban area and opponents focus on the neighbourhoods or blocks with the highest density of Airbnb rentals.

Paying attention to spatial dimensions of tourism also includes taking into account how places and cities are loaded with meaning and values, and how this may distinguish tourists from other visitors, or tourists from residents. Tourist mobilities, perhaps more than any other kind of mobilities, infuse places with meaning. Otherness, authenticity, pleasure, rest, adventure, history, identity, familiarity… all these values may be sought for and performed in tourism, and closely associated to specific places, landscapes, or attractions (see for example Crouch, 2015; Edensor, 2001). Though very diverse, the motivations and expectations for leisure mobilities have generally been, in social sciences, quite clearly distinguished from more constrained mobilities, especially professional trips. In that sense, it is quite problematic to classify indiscriminately, as is the norm in international statistics, business and leisure trips in the same generic category—and to name it "tourism". This should lead us to be particularly attentive to the efforts, in the quantification of tourism, to address the motives of mobilities.

Meanings assigned by tourists to places, indeed, contribute to specific practices of space: sightseeing, strolling, photographing, visiting, shopping for souvenirs… These practices integrate to the vast and diverse set of mingled practices that characterise cities, but may sometimes conflict with more mundane, daily, local practices. Urban geography offers many tools and concepts to examine how different practices and people coexist in cities, and how different appropriations of urban space are embedded in uneven power relations. Studies of gentrification in particular have extensively researched the contribution of tourism to the process (Cocola-Gant, 2018; Gotham, 2005). And in cases where tourism raises "protest and resistance", tourism is "best understood as part of wider struggles around contemporary urban restructuring, the transformation of urban governance and the 'Right to the City'" (Colomb & Novy, 2016: 6). In our case, we will be particularly focusing on the ways the appropriation of these struggles by citizens or local actors relate to the quantification of tourism, be that by putting forward a lived, vernacular knowledge of tourists' presence and figures, or by engaging more directly with numbers by practicing "statactivism" or "map activism" (see for instance Briand, 2017; Cox & Haar, 2020).

1.3. Digital studies

In order to investigate the increased significance of ICT-based data available for the production of tourism statistics, the theoretical framework built for our research should take into

consideration the field of digital studies. In detail, the concepts that can be useful to introduce are those of *Big Data*, *Smart City* and *Platform Urbanism*. Even if there are no exclusive definitions related to these concepts and the phenomena these terms attempt to grasp are in some case intangible, these keywords are widely used in scientific literature and non-technical speeches and a preliminary comment on them is suggested in order to frame our research.

1.3.1. *Big data* and tourism statistics

Big data is a term that has become popular during the last decades, referring to the exponential growth of data produced through the use of electronic devices, especially connected devices. The 3Vs model first formulated by Laney (2001) identifies three core characteristics that set big data apart from classical data: their Volume (the large amount and perceived completeness of data, n=all (Reif and Schmücker, 2020: 2)), their Velocity (the speed of data flow, real-time collection and 'nowcasting' (Song and Liu, 2017: 17)) and their Variety (different formats of data; structured, unstructured and semi-structured (Owais and Hussein, 2016: 254)). Big data have been further classified on the basis of the producer of data (who or what produces the data) and of the production process. For instance, Li et al. suggested that big data can be intended as generated by three sources: users (UGC), devices (sensors), operations (transactions and web activities) (Li et al., 2018: 302); Kitchin proposed a distinction based on directed (e.g. surveillance technology), automated (digital devices recording their use) or volunteered (UGC) data (Kitchin, 2014: 4). A single, shared and unambiguous definition of big data does not exist. Nonetheless, all the elements proposed in these studies can inspire a more complete research and scrutiny of the universe of big data applied to tourism.

Tourists can be counted by the traces they leave during their journeys. In the near past, several methods for tracking the tourists have been of common use. For instance, data on the use of local transports (Terrier, 2006: 57), data from flight booking systems, *in loco* financial transactions (Ostertag and Wöber, 2010: 9-10) and electronic credit cards tracking have been used as statistical sources. However, during the past ten years, the innovations in the field of ICT and the proliferation of the use of mobile phones, smartphones and PCs, have permitted the exploitation of these new sources of data for the statistical analysis of tourism. The real novelty in terms of data production derives by the fact that every kind of online activity performed by tourists, for instance, in social media, booking platforms and web search engines can be collected. The analysis of tourists' inputs (comments, likes, research, reviews…) on online applications such as social media and search engines appears to be increasingly useful for destination marketing organisations (DMOs) to provide a more exhaustive description of tourist practices, sentiments, and consumption patterns (Mariani et al., 2018; Önder et al., 2020). The tourism industry is thought as a pioneer in the use of big data (Demunter, 2017: 6-8; Li et al., 2018: 317-8).

Through the massive use of portable devices, in particular smartphones, people are increasingly leaving digital footprints that can be precisely located. Thus, tourists can be tracked through a wide range of data such as GPS data, mobile networks operators (MNO) data (positioning and roaming data), Bluetooth and Wi-fi events (connections) (Li et al., 2018, p. 310-314). The MNOs have at their disposal mobile positioning data (MPD) deriving from cell phone connections that can be collected even without any active operation of the users (passive MPD) (Reif and Schmücker, 2020: 1) and that are automatically stored in log files by mobile operators (Baggio and Scaglione, 2017: 150). In recent years, passive MPD have been among the most commonly used traces collected with the aim of quantifying tourism flows, allowing for instance to give estimates of the entire present population in a certain place (Terrier, 2006: 60), or specifically of international tourists (Grassini and Dugheri, 2021: 53-54). Studies on the possible use of MPD as a source in official tourism statistics at national level (Ahas et al., 2008; Cousin and Hillaireau, 2019; Saluveer et. al, 2020) and European level (Eurostat, 2014) have shown that through a mixed-data collection, MPD have the potential to strengthen the current surveys carried out on tourist flows and presences. Nonetheless, barriers and discontinuity in data access and

privacy concerns persist as significant problems related to the use of MPD for official tourism statistics (Grassini and Dugheri, 2021: 63).

Domain	Data type (e.g.)
Mobile Communication	MPD, GPS, Wi-Fi, Bluetooth…
Sensors and Wearable Devices	GPS, RFID, NFC, Physiological sensors …
Cameras/Satellites/Photo radars	Images, video recordings…
Business Process Generated Data	Financial Transactions, Credit Cards, Booking engines transactions…
Websites	Searches in search engines…
Social Media	User Generated Content and location data in *Facebook*, *Instagram*…
Online Applications	Comments and reviews

Table 1: Simplified classification of *big data* in Tourism
Based on Reif and Schmücker (2020)

But for tourism as for other domains of quantification, big data have limitations and issues. From the practical point of view, the discontinuity of access, unaligned definitions, selective bias, lack of comparability (Önder et al., 2019: 17-20) and standardisation are recurring issues. Moreover, while big data are often said to be more representative of the 'real world' (Prince, 2020: 2), it is difficult to identify the right data to collect, give them significance, and evaluate the representativity of their population regarding the social phenomenon under study (Song and Liu, 2007: 26; Rich, 2020: 203; Volo, 2020: 307). More generic issues include the problem of harmonisation of the new data processing techniques with the 'classical' ones, since they could be thought as complementary(Baggio, 2019: 263-4); the basic necessity of distinguishing tourists from non-tourists digital traces (Reif and Schmücker, 2020: 2-3) ; security and privacy issues that can result from data abuse and misuse (Song and Liu 26, 2007; Kitchin, 2015); and the risk of neo-positivist approaches that could lead researcher to rely excessively in uncovered correlations without looking for causation processes (Reif and Schmücker, 2020: 3).

1.3.2. Power and governance in the "data-driven" city

As a consequence of the third industrial revolution, the expansion of ICTs appears to have configured a pervasive propagation of software in urban physical spaces that enables the phenomenon of datafication of everyday life (Carta, 2019: 47). It is possible to conceptualise the ecosystem in which city-software works as constituted by (1) the IoT that encompasses all the physical sensors capturing a constant flow of information from, among others, buildings and individuals and (2) the processing of the resulting big data (Mosco, 2019: 60-4). This processing, through the use of algorithms and machine learning (Rich, 2020: 200-1) could help manage urban dynamics with strategies such as crowd control, flow simulations and behaviour recognition (Carta, 2019: 63), and promote informed policies (Kitchin, 2014: 6-7).

The evolution and diffusion of ICTs and the use of big data in urban governance helped generating the concept of *smart city* in the collective discourse. The 'ideal' smart city would be based on the collaboration of different stakeholders (citizens, businesses and public institutions) and systematic integration of ICTs in urban infrastructure to achieve efficiency of urban operation and services and improve the quality of life (Kahn et al., 2017: 1; ECOSOC, 2016: 3-4). The multinational private corporations operating in ICT, driven by technological solutionism and profit interests, seemed to have been the primary and most enthusiast advocates of the application of the idea of smart city. But many governments and public institutions, readily embracing such views,

have contributed to expand the concept of smart city to those of sustainability, inclusivity, and increased quality of life for the citizens (Mosco, 2019: 28-31).

Some scholarly critics however pointed out the limits and flaws of this model. A first criticism focuses on the technical issues deriving from an extensive use of technology in city administration processes. Examples are the issue of the reductionism of complex socio-political and cultural aspects to machine-readable data (technocratic governance based on technological solutionism) and the vulnerability inscribed in the partial and fallible nature of software that makes it prone to security breaches and failures (Kitchin, 2014: 9-11; Mosco, 2019: 224-8). Even if algorithms are perceived as neutral and more efficient than human decisions, such as in the model of *algorithmic governance*, where the management of service is delegated to algorithms (Smith, 2020), these same algorithms are influenced by the political and ideological frameworks surrounding them (Kitchin, 2014: 8) as the code is always built on human (coders') assumptions and categories of understanding (Graham et al., 2019). A second criticism encompasses all the political challenges that the assimilation of technology into public city governance brings about. A major issue arising from the digitalisation of public governance services is represented by the enhanced control systems that seem to permit a new form of mass surveillance (Mosco, 2019: 68-9). This mass surveillance apparatus is embodied by the smart control rooms, which represent the nerve centre of the smart city and collect simultaneously every data from countless sensors, echoing a *panopticon* (Kitchin, 2014: 11-12). The concerns with the smart control rooms are the lack of transparency in their operations (Mosco, 2019: 68-9) and the disruption they bring to power relations (private and public) in the government of cities (Caprotti, 2019: 2476).

But in a more diffuse, less obvious way than control rooms, the private corporations, having the necessary technological *know-how* and products, are progressively penetrating the processes of city governance. Combined with issues of marketisation of public services (Kitchin, 2014: 10) and commercial ownership of data (Mosco, 2019: 222-3), that could produce a loss of democratic control over these data. The term *platform urbanism* designates this emergent reality of digital platform enterprises establishing themselves as major urban service providers, and thus reconfiguring urban materialities and mechanisms, in particular labor, governance, and infrastructure (Leszczynski, 2020). Tourism activities make intensive use of some of these services, in particular accommodation (Airbnb), transport (Uber) or peer-to-peer recommendation (TripAdvisor, Yelp). As mediators between supply and demand, these companies own and operate large databases and can generate, as a by-product of their everyday business, quantitative data relevant to tourist activity. In the perspective of listing the sources and actors of quantification, they must be considered, then, as data producers; but they must be approached bearing in mind their specific relation to data, which is not centred on knowledge as it is with traditional producers of statistics, but directly operational and profit-driven. It is necessary to understand in detail the precise interests or risks for these companies to share or sell their data. They may have different strategies depending on the values datasets have for their different business models: the value of some databases (Airbnb accommodation, TripAdvisor sights, peer-to-peer reviews…) lies in their public availability, while others are selectively displayed by algorithms (Uber vehicles). Besides, many platform companies sell "secondary" datasets or data services, i.e. analytics based on the monitoring of their databases and data flows. There are, thus, several different ways of extracting financial value from platform data (Sadowski, 2019).

It is therefore important to examine the effects these developments have on political balances in cities. As data become an increasingly central resource, the responsibility for regulation of urban services and infrastructure might be progressively transferred to the actors that control the relevant data— this is apparent in the model of "algorithmic regulation" (Ferreri & Sanyal, 2018)—and there is a risk of an increased dependency to particular (monopolistic) platforms, making way for corporate technocratic governance (Ferreri & Sanyal, 2018; Kitchin, 2014).

2. "Mapping" actors and datasets of the quantification of tourism

The first set of methods our research will apply is directed towards a "mapping" of actors and datasets of the quantification of tourism, in order to answer the "descriptive" questions among the research questions—who, what, when and where. By "mapping", we mean not only listing and characterising the main actors and datasets, but also positioning them in relation to each other, and in relation to their role and importance within what we will call "systems" of tourism quantification. The methodological protocol described below will be applied to each case study, countries (France, Italy and Switzerland) as well as cities (Barcelona, Lucerne, Paris and Venice). The aim is to characterise each case's specific tourism quantification system, though different cases can of course share some key actors or datasets.

2.1. Exploration: interviews and documentary collection

The first step is the identification of the main datasets and main actors that have an interest and are involved in the field of tourism statistics for our case studies. It relies on (1) a few semi-structured interviews with people who seem, at first glance, relevant in the institutions that have a central role in the tourism quantification system (those in charge of producing, elaborating, or publishing statistics), and (2) the collection of documents and published information related to the production and producers of tourism statistics such as descriptions of the sources or methods of survey and collection, or mission statements from institutions.

(1) Each of the three countries and the public administrations (municipalities) of the cities selected as case studies rely on specific national, local and regional offices for the collection, production and distribution of tourism data. Therefore, these offices can be considered as a convenient entry point of this exploratory research phase. Within these public institutions we identified the services and persons directly involved in tourism statistics production. We considered interviewing these persons as a necessary step to provide answers to the following preliminary questions leading our research:

- Are there other organisations taking part in the production of tourism statistics at the local/regional/national level?
- What datasets are collected directly by the organisation? What datasets come from other sources? What are these sources?
- Are there any emerging companies or organisations which could challenge the central role of the local/regional/national public institution in the tourism quantification system? Or bring original, alternative, or complementary data on certain aspects of tourism?
- What are the main uses of tourism statistics?

These first interviews can provide a useful outline of the tourism quantification system. Nonetheless, most of the actors interviewed are professionals from public institutions which have a public service mission and should follow specific transparency criteria. While datasets produced outside of this framework, in particular by private polling or market research companies, and the people producing them, have not been taken into consideration in this first stage of the analysis.

(2) The documentary exploration takes place mainly online, starting from the websites of the main institutions involved in tourism statistics and through keyword-based[4] research on search engines. Public statistics organisations are required to publish data widely and show transparency on their methods, and thus provide statements on the missions they are given, definitions of statistical notions and indicators, and explanations of the production processes of statistical datasets. However, these explanations are often inexhaustive and the lists of variables and the granular data are not always available for free. Outside of these public statistics organisations, the wider search of statistical publications allows to complete the research of statistical datasets

[4] Some keywords used were: tourism statistics/tourism presence (+) *name of the city/state*

producers. This allows in particular to explore the field of private companies that produce tourism statistics, as they are often hired by private or public organisations to produce situational or thematic surveys. Tourism businesses are indeed important clients for polling organisations and market research companies. Finally, this documentary exploration provides a general view of the main channels and formats for communicating statistics: databases or data tables, statistics compilation books, regular or one-time analysis reports, etc.

We will go back and forth between points 1 and 2 above once these two approaches have been initiated. That is to say, through the exploratory interviews we will be able to identify the central institutions in the set of actors, as well as the data they disseminate and the publications that serve this purpose. But also, thanks to the analysis of these publications, we will be able to identify "hidden" actors, referenced in these reports, who might have been omitted during our exchanges with our first interviewees. It will be "iteration" (de Sardan, 2013), or the repetition of the same information from one source or another, that will indicate to us that we have completed the exploratory phase. In other words, when the documents available online and the interviews with the first actors identified lead us to the same sources of information, we will have an important clue that the exploratory phase will have been well underway.

2.2. Inventory and mapping

This second step aims to characterise in a more detailed way the actors and datasets, and to classify them in preparation of the critical analysis of the tourism quantification system. This phase of the research is aimed at acknowledging both the specificities of each element, and the categories and relations that organise the system. Descriptive "portraits" (1) aim at grasping the specificities, while relational diagrams (2) outline the comparative and relational analysis of the system's actors.

(1) The redaction of descriptive portraits of actors and datasets is made through the characterisation of some basic elements, but also by adjusting the description to individual specific features. For each actor, we seek to specify, when possible and relevant, the *mission*, the *organisational structure*, the *main datasets produced*, the channels of *publication and access*, and the *place within the tourism quantification system*. Such a description, by relying almost exclusively on the actor's self-presentation, or on the presentation of their own datasets, offers no critical distance, but gives a good idea of the positioning and *raison d'être* of each element. For each dataset, we seek to specify the *type of data* (survey, registry, database…), the main *themes and variables*, the *status* (public, private, open…), and the *format* in which it is available.

(2) Having listed the actors and datasets involved in the tourism quantification system and having gathered detailed information on each one of them, it becomes possible to elaborate a visual representation—a "map"—of a tourism quantification system (see Figure 2). Compared to a textual analysis, such a diagram is inevitably selective and condensed. Indeed, a limited number of parameters can be displayed visually in an efficient manner. Hence, the diagrams aim at characterising and differentiating the elements through criteria viewed as the most essential; and at representing the relations between these elements.

We conceive tourism quantification systems as structured mostly by the actors who produce statistics that are directly relevant to tourism. The diagrams, then, do not represent the users or consumers of these statistics, unless they have a key role in producing other tourism-relevant data themselves, or in aggregating or disseminating tourism statistics.

The complexity of identifying these actors may vary according to their scale of action. In the case of France, for example, municipalities do not have competences recognised by the state with regard to the production of statistics on tourism. Moreover, the municipal level is practically devoid of decentralised or deconcentrated statistical production bodies (Bardet, 2000). As with the centralisation of the countries considered, it is the state and, by delegation, the regions (or prefectures) which have the competence to produce official statistics (although this does not

prevent the appearance of offices dedicated to statistical studies at the municipal level). Therefore, we proceed with the documentary exploration at the local level in by:

- identifying the municipal or metropolitan office in charge of tourism management and promotion,
- analysing their publications on tourism statistics as a basis for a branching work (the reference to a source led us to the supplier of the data in question).

Even data producers may be excluded from the diagrams if we do not judge their role to be pivotal. For instance, actors would be excluded if their data appear only peripheral to tourism, have a very limited audience (measured in particular by the number of other actors who cite them as sources), or are not relevant enough at the considered scale (for instance, for the national scale, museums or parks' frequentation figures). Multiple single elements can be merged into a single relevant category, for instance transportation companies when they all provide data to a central organisation, as is often the case in large cities.

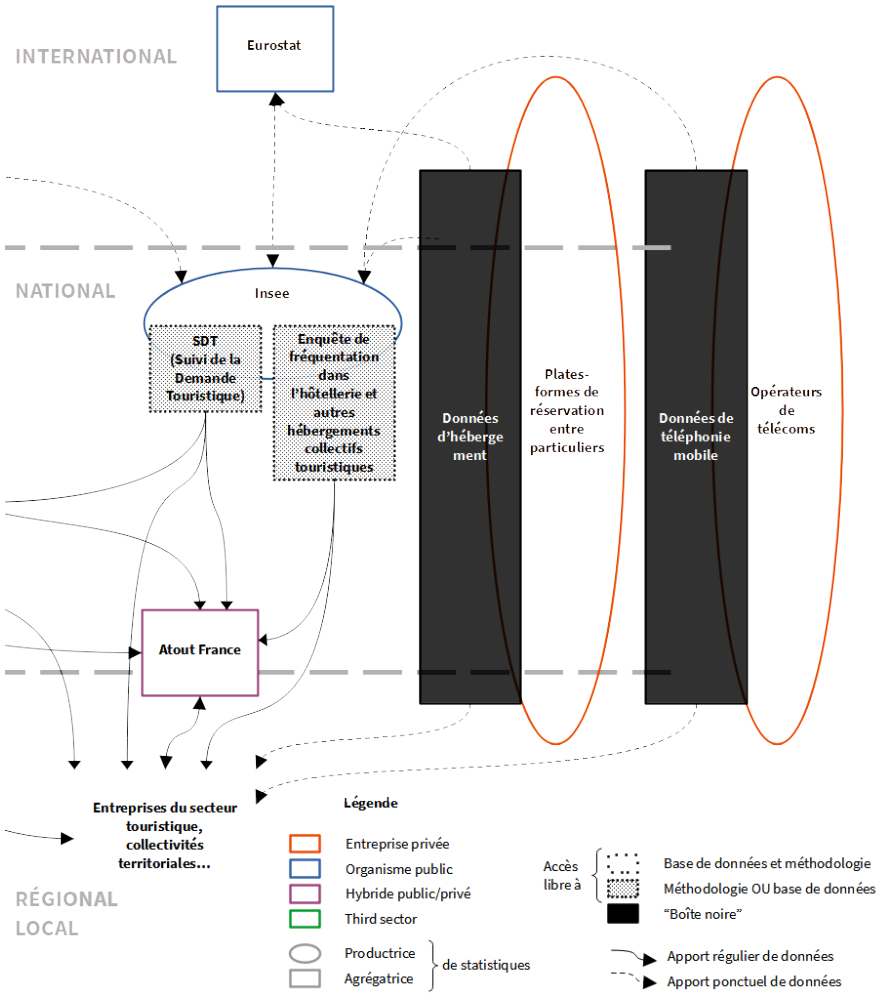

Figure 2: diagram (section) of the French quantification system

Finally, for purposes of clarity and more efficient visualisation, only the most essential or most discriminating features of the actors or the datasets are integrated into the legend. The selected typologies of the elements represented in the map are based on the following parameters:

Status of the organisation: The public or private status of an organisation has many implications in its relation to statistics and data; public and private actors generally have different aims or missions. Public producers of statistics generally have a mission to produce knowledge

and make it accessible to the wider public, in domains that are deemed relevant by governments. On their side, private companies that produce statistics or other data may produce them as a commercial product, or to support their business activities. Nevertherless, public organisations often play a role of support to economic activity, especially in relevant economic sectors such as tourism. Their role may include the facilitation of exchanges or production of knowledge that is directly useful to the private sector. These endeavours often take the form of public-private partnerships or associations—which is why we include the hybrid public/private status in the typology. Finally, the third sector (neither public nor for-profit private organisations) may produce statistics, or aggregate and publish statistics, with specific purposes such as raising awareness on a public issue, assess its the evolution or the effects of their activity on this issue, oppose the action or discourse of another actor, etc.

Determining the status of and actor is pivotal to understand the different positions, interests and finality toward statistics and to represent the distribution of roles between the private, public and third sector within the tourism quantification system. Relations of hierarchy, cooperation, competition or centrality indeed affect the power balance and the governance of the quantification system along public and private interests.

Role in the quantification process: as explained above, the producers of tourism statistics have been considered the key actors in the tourism quantification system. This is reason why they are displayed prominently in the diagrams. The tourism quantification system, defined by all the actors that take part in the prodcution, elaboration, diffusion and general use of statistics is much wider than this display but, for practical reasons, in this first phase of the research only those who have been found to be the most relevant actors are shown.

These actors relate to tourism statistics in disparate ways. The most relevant examples are the actors who "order" the statistics, such as national governments funding a central statistical office or the various organisations which regularly hire poll and survey companies to investigate on specific matters; actors who aggregate or compile statistics from various sources on a given area or issue and publish them and actors who analyse or interpret statistics for their own needs and goals, and more generally all users who discuss and disseminate statistics, in particular with the aim of shaping a discourse on a public issue (such as *overtourism)*. Hence, for mapping purposes we selected the two key roles in the quantification process:

- Producer
- Aggregator

For later analysis, we intend to include the following roles:

- Commissioner, customer, purchaser—those who request tourism statistics
- Intermediary, redistributor
- Analyst, interpreter
- User, consumer

Scale of quantification or action: in the diagram, the actors are positioned according to their main geographical level of action. The transnational companies and the organisations acting toward international statistical coordination are included in the 'International' section. National statistical institutes, national federations of economic actors, ICT's corporations or all other actors able to produce a nation-wide statistical discourse are positioned in the 'National' section. Finally, the many actors producing and aggregating statistics at a regional or local scale to build knowledge of a specific region or city were displayed on the 'Regional and Local' sectors.

As mentioned above, the definition of the relevant area to consider in the making of tourism statistics can be problematic. Regional and local tourism statistics often rely heavily on national and international data, and can also reciprocally fuel upper-level statistics. Even if they do not actually produce data, some regional and actors have a key role in aggregating and interpreting statistics from different sources and different scales—for instance regional or metropolitan

tourism boards. Statistics may even be produced at local or micro-local scales, such as neighbourhoods, streets, or building blocks—such level of precision may be particularly useful to appreciate the phenomenon of *overtourism* and to improve the management of touristic flows. Some actors operate at different scales, for instance platform companies which try to negotiate with upper levels of government (national or European), while also often negotiating with, and even providing data to, city governments.

We attempt to display each actor on a specific section (scale). However, in some cases, it appears necessary to display actors in more than a section. Displaying actors and datasets on different section (at different scales) allows to see the extent of the action of these actors and the way the data circulate between different scales and levels of power.

Openness of the dataset: another major distinction that was integrated in the legend of the diagram is the degree of accessibility and transparency of the key datasets. The level of accessibility/transparency can reflect both (1) the status of the dataset, being it a commercial product or asset or a tool for public knowledge and (2) the possibility to verify the quality of the given dataset (the more opaque, the less verifiable).

Assessing the openness of a dataset is a complex issue. First, the access to granular data, even for public statistics, is often hindered by anonymity or confidentiality obligations. Second, the methodologies used in statistics production are generally complex, and are usually not disclosed to a very detailed level (excluding specific operation of data collection or data processing). Therefore, producers chose to disclose their methodologies are at very different degrees. Quality standards for public statistics have been established, for national or international coordination purposes (e.g. Eurostat protocols), or to ensure the reliability of surveys in the analytics market (e.g. the ISO 20252 certification). While, in the case of the private companies' datasets, these are generally kept 'undisclosed' since they are considered commercial assets. The issue arising from this 'undisclosed' private datasets regards the use some third-party make of these datasets by publishing them to the public for militant purposes (such as in the case of InsideAirbnb). In light of this issue (elaboration of data by third parties), it has been decided to display basic typologies of datasets, and to distinguish them on the basis of accessibility of the granular data and the methodology used.

Finally, the arrows represent the flow (exchange or use) of data, pointing from the producer to the recipient. Regular exchanges of information (established collaborations between organisations) are distinguished from one-time exchanges.

In order to design an efficient visual representation, the typologies of actors and datasets are simplified and some important typologies are even entirely left out of this map. The main features that are not represented in the diagram, due to a high complexity in their visualisation and integration with the other features, but that will be operational in the research are:

- the **type of statistical datasets**: survey, registry, big data/digital footprints (Desrosière's classic distinction, and the question of new sources brought by digitalization),
- the **quantified object/aspect of tourism** in the datasets: tourist presence, accommodation capacity, economic impact, environmental impact, tourists' behaviour.

This last descriptive device should be considered critical since it permits to:

- foster a reflection on the broader dimensions and questions surrounding tourism, by looking beyond the statistical variables to which they are quantitatively reduced.
- question which areas are well covered by existing statistics, and which areas lack a numerical description.
- emphasise the gaps in tourism quantification (e.g. measuring visits to friends and relatives).
- frame the contemporary issues related to *overtourism* and highlight a possible lack of representation of these phenomena in current tourism statistics (e.g. tourist presence outside the traditional accommodation sector).

2.3. Use of the mapping after the exploratory stage

The outcome of this first stage of our research work will, as indicated, take the form of a detailed diagram of the actors in the process of quantifying tourism (see Figure 2), representing the corporate nature of the different entities, the links between them with regard to the circulation of data, their spatial scales of action and the way in which their data is disseminated/published (or not). This leads us to analyse not only the actors who produce statistical data, but also those who commission and use them for various purposes.

This schematic representation of the links between actors and the data they produce will show us clearly the direction of circulation of these data, but also the intersections, dead-ends, and density of information flows between the different stakeholders. In doing so, we will be able to identify the points where central information cannot be obtained through simple documentary exploration. We can imagine, for example, the identification of two private actors involved in the production of statistics on hotel occupancy in Paris, one commissioned by the Region, the other by the Tourist Office. Why "double" this information? Doesn't INSEE already produce disaggregated statistics on this subject? The identification of the actors in charge of the commissioning and production of these studies requires the development of this first mapping resulting from the exploratory stage.

An iteration process will then have to be carried out. The answers of the respondents will adjust what we will have interpreted after the documentary exploration stage: the diagram leads us to the actors and the actors adjust the schema (this work will be done continuously throughout this first stage of our research project, the mapping will constantly evolve). Iteration, that is, the repetition of the information obtained through the first interviews and the first documents consulted, will indicate to us that the exploratory stage is nearing its end (de Sardan, 1995).

Then as detailed in the next section, we will proceed to the in situ surveys. We will spend several weeks in the cities we are interested in, particularly in the institutions responsible for producing/disseminating tourism statistics (tourism offices and/or observatories, deconcentrated offices of public statistics institutes). This is mainly in line with the ethnomethodological approach, since understanding the "production of meaning" of a specific environment requires "the slow and continuous impregnation of the human groups with which we maintain relations" (Laplantine, 2011: 13). The structure of the interviews and some details on how we plan to develop our ethnomethodological approach will be outlined in the following points.

3. Questioning actors, datasets, and discourses

As mentioned in the introduction of this paper, this phase of the project is designed to answer the last two, more analytical questions of the study of quantification: *how* and *why* are tourism statistics made? But the methods deployed in the field of the quantification of tourism also allow to refine and develop the other questions, *who makes tourism statistics*, *what are these statistics*, *where* and *when* are they produced. Indeed, while the mapping phase relies mostly on secondary material produced by the statistics producers themselves, the field research phase will consist of elaborating our own, primary research material, by interviewing actors, gathering first-hand observations, and getting the closest possible to the processes of making statistics.

The field research will be grounded in the sociology of quantification in that we want to outline the contexts (socio-professional, economic, political, geographical and historical, etc.) in which tourism statistics are produced and the influence of these contexts—and this is one of the reasons why we use multiple study cases. The research is informed by a field study of tourism quantification as a socio-professional context, and aims at analysing the mechanisms of authority, and of contestation, that structure and agitate this socio-professional context. The methods inspiring this research will be borrowed from the classical, critical perspective of social sciences and their skillset: semi-structured interviews, sociological characterisation of actors, discourse

analysis. Nonetheless, the field research will also borrow from the ethnomethodological perspective and its application to statistics-producing work, ethnostatistics. The applied methods will be ethnomethodological since the research aims at observing the ordinary, day-to-day, taken-for-granted practices and preoccupations of statistics producers. The focus of the analysis will be on their *methodical* practices, aimed at producing shared social orders. Moreover, sensemaking and ordering practices of "non-expert" groups, that is to say the practical knowledge ordinary citizens develop of tourism and its quantification, will be included as a subject of investigation. However, as mentioned in the previous paragraph, the perspective on these practices will not be restricted to an internal one, a stance that, as ethnomethodology prescribes, demands a suspension of the critical lens and main concepts of social sciences on the methods under study (Laurier & Bodden, 2020). Finally, it should be noticed that ethnostatistics would not encompass the entire scope of methods utilised, since the field of study will not be limited to statistics producers, but it will also include aggregators and some key users of tourism statistics, as well as datasets that may not entirely meet the definition criteria of 'statistics'.

Finally, the research will take into account different types of groups acting at different levels—the "ethnos" in ethnomethodology. Producers of statistics will be one of the main groups; but within this group they may distinguished, for instance, public and private sector statisticians. Other groups will be less clearly defined as they are not professional groups, for instance the civic associations. Such groups will be delimited by their interests or position within the tourism quantification system or tourism debate rather than by their shared methodical practices.

In sum, the field research will employ a mix of general qualitative social science methods and of ethnomethodology- and ethnostatistics-informed methods. These methods are explained in the paragraphs below.

3.1. Semi-structured interviews

Like our project, our interviews will be designed and organized by scale. It is obvious that specific questions will be asked depending on the person being interviewed and the institution or company to which he or she is attached. It is not our intention here to list all the questions we may ask to each institution and person. Nevertheless, we believe it is relevant to list the general points to be addressed regardless of the scalar nature of our respective case studies. The interview guide below therefore details the points to be addressed in a general way.

The questions have been elaborated around 4 key points: 1. the role of the actor or "work relationship with tourism statistics" 2. the specifics of quantified objects and quantification methods, 3. their relationship with the other actors in tourism quantification, and 4. their perception of the effects that their work produces in the governance and evolution of tourism. These points encompass the ethnomethodological and sociological dimensions of our research. The more geographical considerations will depend on the scale on which we will work according to the respective case studies and fields.

The interviews of the first exploratory phase will also be the opportunity to start our ethnomethodological observations. According to Stéphane Beaud, the interview should not be conceived as a clinical intervention with a patient who must provide us with very precise data in order to establish a useful diagnosis for our investigation. Of course, we aim to obtain specific information through the register of the respondents' words, but the objective of our interviews is also to take advantage of the interview situation, seen as "a scene of social observation" that allows us to better understand the places and the people (Beaud, 1996: 235). For the moment, these "situations of inquiry" are elusive.

The mobilisation of an interview guide does not pretend to make our encounters routine or to push the quantitative analysis through the accumulation of similar opinions and comparable answers. This tool serves to coordinate our efforts and to guide our reflections, once we are in our different fieldwork sites.

Interview guide

1. Work relationship with tourism statistics

- What is the work of your organisation in relation to tourism statistics? (What other persons or departments also work with tourism statistics in your organisation? How many, how are they organised?) What is your personal role in this work?
- How did your professional trajectory (including education and training) lead you to this position?
- What are the aspects and dimensions of tourism that you need quantified, and why? Are there any data that you lack?
- Why is it important, in your opinion, to measure these phenomena quantitatively?

2. Quantified objects and quantification methods
(2.a. or 2.b. depending on the role of the interviewee)
2.a. Producer

- How do you define your main statistical indicators on the tourism phenomenon, and how were they selected and defined at first?
- What are the collection methods of these statistical data?
  (Here, ask precise questions on sources and methods, address the specificities of local or institutional contexts, etc. Try to make the clearest possible the undisclosed assumptions and technicalities of the methodology. Including problems, obstacles, insufficiencies…)
- What are the main processing methods and analyses?
  (Idem)
- Have your work processes changed over time?
- Have technological evolutions changed the collection, processing, or publication of these statistics?
- Are there any possible improvements of these statistics?

2.b. Other role

- What are the main sources or statistical indicators on the tourism phenomenon that you are working with?
- What are the main operations you perform with these statistics? Statistical operations, compiling, cross-referencing sources, interpretations, syntheses…?
  (Here, ask precise questions on sources and methods, address the specificities of local or institutional contexts, etc. Try to make the clearest possible the undisclosed assumptions and technicalities of the methodology. Including problems, obstacles, insufficiencies…)
- Has your work with these statistics changed over time?
- Have technological evolutions changed this work?
- Do you think you can contribute to the improvement of these statistics?

3. Relationships with other actors of tourism quantification

- What are the main clients/providers/users/partners that you exchange tourism statistics with? To what end?
- Are there any organisations doing similar work to yours with tourism statistics? Are they your competitors?
- If you provide data or analyses to other organisations, what main uses do they have of these data?

4. Interpretation on the effect of the data created and disseminated.

- Do you publish any statistical data or analyses? If you do, through what channels or procedures?

- What sectors of the population or the economy may be interested in or affected by these data or analyses?
- Are your data or analyses used in decision-making (public policies, economic sectors…)?
- Can your data or analyses weigh in contemporary debates or challenges regarding tourism? How so?

3.2. Ethnomethodology and ethnostatistics

The main methodological approach inspiring our research is related to ethnostatistics, a specific methodology designed by Gephart (1988) and inspired by the practice of ethnomethodology developed by Garfinkel (Gephart, 1988: 11-13 and Gephart, 2012: 79). In order to frame the ethnostatistical method, it can be useful to introduce the ethnomethodological system, of which ethnostatistics represent a sub-field.

In line with ethnographic research, ethnomethodology could be functional to understand the cultural and social aspects that phenomena and things acquire within human groups. Both research approaches are based on the activities of fieldwork, (participant) observation and interviews. Nonetheless, ethnomethodology supposedly differs from ethnography by the level of analysis. In fact, examining processes within interpersonal actions (e.g. scrutiny of conversations between colleagues), the ethnomethodologist focuses on a finer granularity than the ethnographer (Winiecki, 2008: 191) and moreover does not generally use a definite method (Laurier, 2009). The main focus of the ethnomethodological approach is the common-sensemaking practices (*methods*) performed by the members of a specific group (*ethnos*) to produce shared meanings within this specific group. These methods encompass the practical actions, reasoning, conversations and the day-by-day production of order carried out by people interacting in a specific context (Laurier, 2009). The common-sensemaking through interaction should in fact be considered contingent to the context, or occasion, as displayed by the same members participating in it (Hak, 1995: 113-5).

As noticed previously, ethnostatistics as envisioned by Gephart (1988, 2006) is a specific ethnomethodological approach. The sensemaking practices and interpretative procedures studied by ethnostatistics are those of statisticians, who create shared valuable, reliable and legitimate numerical measures of world phenomena and it aims at understanding the production of social phenomena through their measurement (Gephart, 2012: 79-80).

The ethnomethodological approach aims at an endogenous, internal account of methods and order under study. Hence, it demands acquaintance with the practices of the researched group or phenomenon—ideally, the researcher would become a member of the group. Utilising the ethnostatistical approach brings the advantage of informing the researcher on the 'life cycle of statistics' by permitting a deep observation of the social phenomena and contexts surrounding and determining statistics during 'mundane everyday practice' (Gephart, 1988: 10). Therefore, this approach develops through a critical investigation of the 'untold' non-objective, non-scientifically codified and taken-for-granted practices and inexplicit assumptions of statistics-makers and users. (Winiecki, 2008: 186-8). The configuration of ethnostatistics as a critical tool responds to a need of de-mythifying the 'allure of numbers' and a perceived superiority of quantitative methods over the qualitative ones (Gephart, 1988: 10) that permeates not only research but the 'scientist' Western society *in toto* (Kleinman, 2005; in Winiecki, 2008: 186) that relies on numbers to make reality intelligible, measurable (Gephart, 2012: 74-6) and accepts positivist assumptions and quantitative analysis as objective facts. (Bogdan and Ksander, 1980: 302). As a matter of fact, ethnostatistical research brought to light the facts that technical (quantification) processes rarely produce 'objectively real outcomes' and that they correspond to different 'truths' (Winiecki, 2008: 190).

On the other side, this in-depth qualitative analysis should be envisioned as a tool to refine and improve the validity of measurement methods and therefore of the data produced (Stoycheva and Favero, 2020: 2). Ethnostatistics, in fact, originates in the critical analysis of the objectivity and

'reality' of numbers and statistical artefacts, but its finality is to ameliorate the production of statistics themselves (Gephart, 1988).

Ethnostatistics research develops through the scrutiny of three stages (areas): the production of statistics (quantification and datafication of phenomena), statistics at work (categorisation and analysis of data), and statistics as rhetorics (making data influential in society) (Gephart, 1988; Winiecki, 2008: 204).

1st order: Producing a Statistic

Following an ethnographical approach, the first order ethnostatistics analysis aims at describing the working place, the organisational culture, setting and needs, the symbolic apparatus and categories mobilised by statistics makers and users in their explanation of phenomena, the norms legitimising their (quantifying) activity and the work constraints (Knorr-Cetina, 1983: 134). On a more meticulous level (ethnomethodology), first order ethnostatistics should assist the research in uncovering the individual interests, motivations, concerns and the discursive interactions (Knorr-Cetina, 1983: 128) and choices made among professionals in everyday life that influence the production of statistics (Gephart, 1988: 15-29). This perspective on statistics production derives from a theoretical position that conceives the exercise of science (in this case, quantification of phenomena) as a constructive process rather than a pure and unmediated description of reality. Given that, for instance, even the source material to be enumerated depends on choices, first order ethnostatistics indicates that objectification (fabrication of axiomatic data) of the subjective (data approached and elaborated by statistics makers and users) is made through contingent selections and negotiations (Knorr-Cetina, 1983: 122-124) depending on particular circumstances and agents.

2nd order: Statistics at work

Second order ethnostatistics requires an involvement of the ethnostatistical researcher in the process of statistics making. Its objective is to directly experiment with technical and practical assumptions by elaborating statistics following alternative selections (Berkeley study quoted in Knorr-Cetina, 1983: 121) and methodological assumptions (Gephart, 1988: 30-31) such as *ad hoc* approximations and metrics (Gephart, 1988: 31-35) and by comparing the results of these alternative selections with the originally proposed ones in order to demonstrate that some degree of accidental distortion of data occurs during the process of elaboration.

3rd order: Statistics as rhetoric

Third order ethnostatistics focuses on the rhetorical disposition of statistics. Gephart suggested that pieces of communication that display statistics such as reports should be examined as dramatic narratives emanating persuasive power. Therefore, in order to reveal the practice of rhetoric, these displays should be observed through the lenses of literary criticism (Gephart, 1988: 14) and the analysis of language and graphical representations (Beattie and Jones, 1992: 301).

4th order: Historical ethnostatistics

In addition to the three orders indicated by Gephart, Stoycheva and Favero suggested to refine the research method by taking into account how temporality affects the production, use and transmission of statistics. In detail, through the practice of ethnographic history (Rowlinson et al. 2014) the aim of "fourth order" ethnostatistics is to identify the temporal evolution of the meaning assigned to methods and data and to recognise what is lost and acquired in the transmission of these data (Stoycheva and Favero, 2020: 12-13). The available historical documents used as reference for quantification are the main object of the analysis (including, for instance, drafts and private correspondence) and, since these sources are not constructed but found, they must be interpreted through the lenses of critical realism (Stoycheva and Favero, 2020: 10).

On the basis of the previous theoretical introduction to ethnostatistics, in the next table are summarised the effective steps that are going to be taken during the field research.

Order	Ethnostatistics applied to Research Activity – Tourism Statistics
1st	Conducting ethnographic studies of groups that routinely produce tourism statistics.
2nd	Observing and investigating the function of technical and operational assumptions that are involved in the production of statistics. (Terminology and parameters).
3rd	Examining the use of statistics as rhetorical device in various kind of research publications. (Analysis of language and words)
4th	Reconstructing history (social life) of data.

Table 1: **Simplification of ethnostatistics research methodology**
Based on Gephart's work, as proposed by Stoycheva and Favero (2020: 11).

Thus, our ethnostatistical researcher should ideally focus on:

- Describing the context and investigating the socio-cultural, organisational, individual and temporal factors influencing the process of statistics-making and the theoretical and technical assumptions underlying this process;
- Going 'native' and 'making trouble' (Gephart, 1988:30-31) by simulating statistics-making and testing alternative methodologies of data collection and production. Therefore, whenever possible, collaborating in the editing of tourism statistics documents;
- Encouraging the improvement of methods, processes and data quality though this critical approach to statistics-making;
- Detecting the persuasive features of statistics by the analysis of discourses and publications;
- Analysing and reconstructing the flow of data and data-making knowledge both temporally (transmissions of data, knowledge and practices within the same organisation/group) and spatially (transmission of data and knowledge among different actors).

In our case, with limited skills in statistics, we will not be able to develop an expert understanding of the mathematical stakes of the statistic-making processes—that is to say, second-order ethnostatistics. Testing or suggesting alternative mathematical models will not be within our reach. We will, however, get the most familiar possible with accessible statistical datasets. This implies a detailed understanding of variables, and basic handling of the datasets. Such ethnostatistical methods, which require close observation of work practices, will probably be possible within public statistics organisations, due to their transparency and accessibility standards. Other key actors, whose statistical production is not a core mission, and/or which maintain and process databases as a for-profit activity, will likely be harder to approach with ethnostatistical methods. Since our research project aims at analysing different actors, short periods of direct and participant observation within the working context of some key actors will be undertaken for each case study. The participant observation will ideally take place at both national and at regional/local level in order to produce insights on statistics-production in offices at different level. If close observation of their work practices proves difficult, other methods, in particular interviews and discourse analysis, will provide the main material for these actors.

3.3. Participant observation in urban tourism space

Protocols of participant observation will be undertaken within urban space to apprehend inhabitants' experiences of tourism—and in some cases of *overtourism*. We will consider, in an ethnomethodology-inspired perspective, that city inhabitants have methods to produce meaning and order related to the tourism phenomenon, and even related to the quantification of tourism.

Through *in situ* observation of tourism practices, and informal conversations with local actors such as residents, tourism workers and shopkeepers, we will try to apprehend ordinary encounters with tourists. Indeed, opinions on tourism and ways of acting with tourists are built in everyday interactions (Eggli, 2021). Inhabitants also elaborate on a daily basis a geographical and sociological understanding of their urban environment (Strebel, 2014). Hence, we will investigate to what extent local inhabitants, through embodied and situated assessments of the volume and variation of tourism frequentation, through the observation of material clues and traces (signs of tourist-oriented accommodation, shops, services…) can develop a sense of tourism quantification.

3.4. Analysing discourses on statistics

An exploration of the rhetorical power and uses of statistics is particularly important in our case, as the project addresses current debates on tourism, but also questions potential evolutions of tourism statistics. All of the following themes, which are fueled by tourism statistics or deal with tourism statistics, will be among our major *foci*: the overtourism debate, where statistics can be used to support or perpetuate pro-growth stances, or to defend anti-tourism positions, that we will attempt to analyse and map borrowing tools from controversy mapping (Venturini & Munk, 2021); the criticism of tourism statistics, that we mentioned in academic publications, but that may also be voiced in professional tourism or statistical sectors; the general perspectives on tourism (is the economic perspective really hegemonic? How are shifts in practices or economic trends usually framed?); and innovation, disruption, or datafication talk, as digital economy actors strive to expand their markets, and traditional statistical producers strive to keep up with new sources of data and new methods of data processing.

The analysis of these discourses will be mostly a qualitative data analysis. A corpus, or different *corpora*, of discursive material will be elaborated from various sources. Among the most readily accessible material are the publications on which we also largely base our mapping of actors and datasets (see section 2). Key statistics producers and aggregators, indeed, publish statistical reports that are mostly tables and graphs, but also statistical analyses, where statistics are selected, commented and interpreted. Although the authors of these analyses generally try to adopt a neutral and descriptive stance, their texts are inevitably selective and constructed within professional or political positionalities. Public policy statements or reports published by governments, be they local or national, make another important material, as they may contain recommendations on future tourism developments, public issues analyses, etc. Private companies communicate through marketing campaigns, annual reports to shareholders, public relations and press declarations. These corporate sources can be especially useful to apprehend these companies' positions within controversies raised by their activities—Airbnb in particular has had to engage in many negotiations and discursive efforts to defend their activity against projects of regulation (Ferreri & Sanyal, 2018). Finally, citizen associations and NGOs may also publish or communicate statistics, be they official, professional or stemming from citizen "statactivism", within brochures, websites, media declarations, or even demonstrations.

Conclusion and pending questions

In this document, we presented a research design dedicated to the analysis of the tourism quantification system at different geographical levels and on different case studies. Our approach is rooted in the sociology of quantification, the geography of tourism, and digital studies. It blends efforts of deconstruction of statistics with an analysis of contemporary issues pertaining to urban tourism—issues that are related to manners of counting tourists. The research protocol is ambitious in that it aims at giving a full—though not exhaustive— picture of actors and datasets involved in the quantification of tourism in the selected case studies, as well as a detailed analysis

of how statistics and data are mobilised in current tourism evolutions and debates. To meet these general goals, we laid out a number of tools, and a basic conceptual and practical vocabulary. But how exactly will we put these to good use in our research work? Many further "methods" and ways of interpretation will have to be elaborated according to empirical material and practical constraints, improvised, and adapted in the course of research. Below are a few of the pending questions that this initial research methodology could not fully answer.

How to analyse a statistical "programme"? A statistical project, a use of statistics? More broadly, **how to characterise different relations to statistics or "practices" of statistics?** The research protocol provides a general typology to distinguish these different relations, and ethnostatistical methods allow to describe them in detail. But we still need to reflect on how wecan render or "narrate" the flows of data, the material organisation and work practices that amount to a specific relation to statistics—as producer, aggregator, power user, etc.

How to qualify different datasets according to criteria of opennessand transparency? We emphasised the wide variety of situations in terms of the level of detail of accessible methodology, in terms of actual access to granular datasets or aggregated data. But how to integrate these practicalities with general issues such as individual rights and the commodification of data? The concepts of openness and transparency are morally charged, and they are influential in political debates on the social role of data. Hence, we need to integrate them to the analysis, but to handle them with care. This means that our assessment of datasets needs to follow a very explicit set of criteria.

How do we research the action and role of digital companies and big datasets? Within the quantification of tourism and the production of urban space? How can we approach the strategic role of their databases and algorithms, given that these assets are protected when they constitute a competitive advantage? And how to precisely measure the competition these actors represent to traditional actors and statistics, when the nature and scope of their datasets are so different from the missions and uses of traditional statistics?

How to do an analysis of classifications and statistical variables? How to study commensuration, in particular in the international standardisation of statistical categories? How to assess their adequacy to the tourism phenomenon? This involves a reflection on the "best" definition of tourism—from what perspective? Social sciences, public policy, international organisations?

References

Ahas, R., Aasa, A., Roose, A., Mark, Ü., & Silm, S. (2008). Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management, 29(3), 469-486.

Alonso, W., & Starr, P. (Eds.). (1987). *The Politics of Numbers*. Russell Sage Foundation.

Anderson, M. J. (2015). *The American Census* (Second Edition). Yale University Press.

Armatte, M., (2003). La introducción en Francia de los métodos de sondeo aleatorio. *Journal de la Société Française de Statistique,* 144 (1-2), 227-255.

Baggio, R. (2019). Measuring Tourism : Methods, Indicators, and Needs. In E. Fayos-Solà & C. Cooper (Eds.), *The Future of Tourism : Innovation and Sustainability* (p. 255-269). Springer International Publishing. https://doi.org/10.1007/978-3-319-89941-1_13

Baggio, R., & Scaglione, M. (2017). Strategic visitor flows (SVF) analysis using mobile data. In *Information and communication technologies in tourism 2017*, 145-157. Springer, Cham.

Bardet, F. (2000). *La statistique au miroir de la région. Éléments pour une sociologie historique des institutions régionales du chiffre en France depuis 1940* [Thèse de doctorat en science politique, Paris 1 - Panthéon-Sorbonne]. https://halshs.archives-ouvertes.fr/tel-03101587

Bardet, F. (2012). Distinguer les variables des objets dans la quantification sociologique. Ou comment l'opposition française aux « statistiques ethniques » n'empêche pas de penser la société multiculturelle. *Études interculturelles*, 5(1), 37-49.

Bauder, M., Freytag, T., & Gérardot, M. (2014). Exploring tourist mobility in Paris. A combined visitor survey and GPS tracking study. *EspacesTemps.net*. https://www.espacestemps.net/en/articles/analyser-les-mobilites-touristiques-a-paris-en-combinant-enquete-visiteurs-et-gps/

Beattie, V., & Jones, M. J. (1992). The use and abuse of graphs in annual reports: theoretical framework and empirical study. *Accounting and business research*, 22(88), 291-303.

Beaud, S., (1996). L'usage de l'entretien en sciences sociales : plaidoyer pour l'entretien ethnographique. *Politix,* 35(9), 226-257.

Bogdan, R., & Ksander, M. (1980). Policy data as a social process: A qualitative approach to quantitative data. *Human Organization*, 302-309.

Briand, F. (2017). *Les meublés touristiques à Paris en 2016—Des habitants et une économie sacrifiés au profit d'AirBnB*. https://parisvsbnb.fr/donnees-paris-2016/etude-pdf-sur-limpact-du-developpement-des-logements-airbnb-a-paris-en-2016/

Camarillo-Naranjo, J.-M., Vallejo-Villalta, I., Fernández-Tabales, A., & Santos-Pavón, E. (2022). Where is tourist housing actually located? New approaches and sources for detailed scale analysis. *European Planning Studies*, 30(4), 744-768. https://doi.org/10.1080/09654313.2021.2002825

Caprotti, F. (2019). Spaces of visibility in the smart city: Flagship urban spaces and the smart urban imaginary. *Urban Studies*, 56(12), 2465-2479.

Carta, S. (2019). Big data, code and the discrete city: Shaping public realms. Routledge.

Cocola-Gant, A. (2018). Tourism gentrification. In L. Lees & M. Phillips (Eds.), *Handbook of Gentrification Studies* (p. 281-293). Elgar. https://www.elgaronline.com/view/edcoll/9781785361739/9781785361739.00028.xml

Colomb, C., & Novy, J. (Eds.). (2016). *Protest and resistance in the tourist city*. Routledge.

Cousin, G., & Hillaireau, F. (2019). Can Mobile Phone Data Improve the Measurement of International Tourism in France? *Economie et Statistique / Economics and Statistics*, 505-506, 89-107. https://doi.org/10.24187/ecostat.2018.505d.1967.

Cox, M., & Haar, K. (2020). *Platform Failures—How short-term rental platforms like Airbnb fail to cooperate with cities and the need for strong regulations to protect housing*. The Left in the European Parliament. http://insideairbnb.com/reports/Platform-Failures-FINAL-VERSION.pdf

Crouch, D. (2015). Unravelling Space and Landscape in Leisure's Identities. In S. Gammon & S. Elkington (Eds.), *Landscapes of Leisure—Space, Place and Identities* (p. 8-23). Palgrave Macmillan. http://www.palgraveconnect.com/doifinder/10.1057/9781137428530

Cussó, R. (2003). Les statistiques de l'éducation de l'UNESCO : restructuration et changement politique. *Éducation et sociétés*, 12(2), 57-72.

Daniel, C., (2016). La sociología de las estadísticas. Aportes y enfoques recientes. *Cultura y ciencias, 7 (2016)*, CONICET. Buenos Aires, 73-95.

De Sardan, J-P., (1995). La Politique du Terrain, sur la production des données en anthropologie. *Revue Enquête. Les terrains de l'enquête,* 1, 71-109, EHESS/Parenthèses.

Demunter, C., (2017) 'Tourism statistics: early adopters of big data', *EUROSTAT*, Luxembourg.

Desrosières, A., (2010). La politique des grands nombres. Histoire de la raison statistique. La Découverte. Paris.

Desrosières, A., (2014). Prouver et gouverner. Une analyse politique des statistiques publiques. La Découverte. Paris.

Diebold, F. X. (2012). On the Origin(s) and Development of the Term 'Big Data'. PIER Working Paper No. 12-037. https://ssrn.com/abstract=2152421

Dijkstra, L., & Poelman, H. (2014). A harmonised definition of cities and rural areas : The new degree of urbanisation. Working Papers, European Commission.

Duchêne-Lacroix, C., Hilti, N., & Schad, H. (2013). L'habiter multilocal : Discussion d'un concept émergent et aperçu de sa traduction empirique en Suisse. *Quetelet Journal*, 1(1), 63-89. https://doi.org/10.14428/rqj2013.01.01.05

ECOSOC (2016) Commission on Science and Technology for Development, 'Smart cities and infrastructure', Nineteenth session Geneva, 9–13 May 2016 Item 3(a) of the provisional agenda.

Edensor, T. (2001). Performing tourism, staging tourism : (Re)producing tourist space and practice. *Tourist Studies*, 1(1), 59-81. https://doi.org/10.1177/146879760100100104

Eggli, F. (2021). Living with tourism in Lucerne. How people inhabit a tourist place [PhD thesis]. University of Lausanne.

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a Social Process. *Annual Review of Sociology*, *24*(1), 313-343. https://doi.org/10.1146/annurev.soc.24.1.313

Espeland, W. N., & Stevens, M. L. (2008). A Sociology of Quantification. *European Journal of Sociology / Archives Européennes de Sociologie*, *49*(3), 401-436. https://doi.org/10.1017/S0003975609000150

Ferreri, M., & Sanyal, R. (2018). Platform economies and urban planning : Airbnb and regulated deregulation in London. *Urban Studies*, *55*(15), 3353-3368. https://doi.org/10.1177/0042098017751982.

Gephart Jr, R. P. (2006). Ethnostatistics and organizational research methodologies: An introduction. *Organizational Research Methods*, *9*(4), 417-431.

Gephart, R. (2012). Ranking Research: Toward an Ethnostatistical Perspective on Performance Metrics in Higher Education. *Recherches en Sciences de Gestion*, 93, 73-90.

Gephart, R. P. (1988). *Ethnostatistics: Qualitative foundations for quantitative research* (No. 12). SAGE, London.

Gotham, K. F. (2005). Tourism Gentrification : The Case of New Orleans' Vieux Carre (French Quarter). *Urban Studies*, *42*(7), 1099-1121. https://doi.org/10.1080/00420980500120881

Graham, M., Kitchin, R., Mattern, S., & Shaw, J. (Eds.). (2019). *How To Run A City Like Amazon, And Other Fables*. Meatspace Press. https://meatspacepress.com/how-to-run-a-company-like-amazon-and-other-fables/

Grassini, L., & Dugheri, G. (2021). Mobile phone data and tourism statistics: a broken promise?. *National Accounting Review*, 3, 50-68.

Hacking, I. (2008). *Entre science et réalité : La construction sociale de quoi ?* La Découverte.

Hak, T. (1995). Ethnomethodology and the institutional context. *Human Studies*, 18(2), 109-137.

Hall, C. M. (2005). Reconsidering the Geography of Tourism and Contemporary Mobility. *Geographical Research*, *43*(2), 125-139. https://doi.org/10.1111/j.1745-5871.2005.00308.x

Jugnot, S. (2013). *Catégoriser les origines pour outiller la connaissance ou pour outiller l'action ? : Quelques enseignements de l'exemple canadien*. Journées d'étude sur les données longitudinales dans l'analyse du marché du travail. 25-26 juin 2013, Centre Emile-Durkheim, Bordeaux. 29-41.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big data & society*, *1*(1), 2053951714528481.

Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal*, 79(1), 1-14.

Knorr-Cetina, K. D. (1983). The ethnographic study of scientific work: Towards a constructivist interpretation of science. Sage, London.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META group research note*, 6(70), 1.

Laplantine, F., (2011). *La description ethnographique*. Armand Colin, Paris.

Laurier, E., & Bodden, S. (2020). Ethnomethodology/Ethnomethodological Geography. In A. Kobayashi (Ed.), *International Encyclopedia of Human Geography (Second Edition)* (p. 329-334). Elsevier. https://doi.org/10.1016/B978-0-08-102295-5.10261-6

Leszczynski, A. (2020). Glitchy vignettes of platform urbanism. *Environment and Planning: Society and Space*, *38*(2), 189-208. https://doi.org/10.1177/0263775819878721

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research : A literature review. *Tourism Management*, *68*, 301-323. https://doi.org/10.1016/j.tourman.2018.03.009

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism : A systematic literature review. *International Journal of Contemporary Hospitality Management*, 30(12), 3514-3554. https://doi.org/10.1108/IJCHM-07-2017-0461

Mosco, V. (2019). *The smart city in a digital world*. Emerald Group Publishing.

Önder, I., Gunter, U., & Gindl, S. (2020). Utilizing Facebook statistics in tourism demand modeling and destination marketing. *Journal of Travel Research*, 59(2), 195-208.

Ostertag, J. and Wöber, K. W. (2010). European city tourism statistics. In J. A. Mazanec & K. W. Wöber (Eds.), *Analysing international city tourism* (p. 25-41). Springer, Vienna.

Owais, S. S., & Hussein, N. S. (2016). Extract five categories CPIVW from the 9V's characteristics of the big data. *International Journal of Advanced Computer Science and Applications*, 7(3), 254-258.

Pistre, P., & Richard, F. (2018, April 27th). Seulement 5 ou 15 % de ruraux en France métropolitaine ? Les malentendus du zonage en aires urbaines. *Géoconfluences*. http://geoconfluences.ens-lyon.fr/informations-scientifiques/dossiers-regionaux/france-espaces-ruraux-periurbains/articles-scientifiques/definition-espace-rural-france

Porter, T. M. (1985). The Mathematics of Society : Variation and Error in Quetelet's Statistics. *The British Journal for the History of Science*, 18(1), 51-69. https://doi.org/10.1017/S0007087400021695

Pratt, S., & Tolkach, D. (2018). The politics of tourism statistics. *International Journal of Tourism Research*, *20*(3), 299-307. https://doi.org/10.1002/jtr.2181

Prince, R. (2020). The geography of statistics: Social statistics from moral science to big data. *Progress in Human Geography*, 44(6), 1047-1065.

Rawls, A. W. (2008). Harold Garfinkel, Ethnomethodology and Workplace Studies. *Organization Studies*, *29*(5), 701-732. https://doi.org/10.1177/0170840608088768

Reif, J., & Schmücker, D. (2020). Exploring new ways of visitor tracking using big data sources: Opportunities and limits of passive mobile data for tourism. *Journal of Destination Marketing & Management*, 18, 100481.

Rich, R. (2020). 'Big, Thick, Small and Short-The Flaws of Current Urban Big Data Trends', *Geography Research Forum*, vol. 40, 193-206.

Sadowski, J. (2019). When data is capital : Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 205395171882054. https://doi.org/10.1177/2053951718820549

Saluveer, E., Raun, J., Tiru, M., Altin, L., Kroon, J., Snitsarenko, T., Aasa, A., & Silm, S. (2020). Methodological framework for producing national tourism statistics from mobile positioning data. *Annals of Tourism Research*, 81, 102895.

Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research : The first decade. *Tourism Geographies*, 18(5), 587-606. https://doi.org/10.1080/14616688.2016.1214977

Smith, G. J. (2020). The politics of algorithmic governance in the black box city. *Big Data & Society*, 7(2), 2053951720933989. https://doi.org/10.1177/2053951720933989

Song, H., & Liu, H. (2017). Predicting tourist demand using big data. In Z. Xiang & D. R. Fesenmaier (Eds.) (p. 13-29). *Analytics in smart tourism design*. Springer, Cham.

Stock, M., Coëffé, V., & Violier, P. (2017). *Les enjeux contemporains du tourisme : Une approche géographique*. Presses Universitaires de Rennes.

Stoycheva, S., & Favero, G. (2020). Research methodology for ethnostatistics in organization studies: towards a historical ethnostatistics. *Journal of Organizational Ethnography*, 9(3), 327-342. https://www.doi.org/10.1108/JOE-03-2019-0016

Strebel, I. (2014). Re-specifying geographical quantification : Problems of order in street interviews. *Transactions of the Institute of British Geographers*, 39(2), 278-290. https://doi.org/10.1111/tran.12014

Terrier, C., Armand, L., Antczak, M., Khiati, A., & Sylvander, M. (2007). Mobilité touristique et population présente : Les bases de l'économie présentielle des départements. Ministère des transports et de l'équipement, du tourisme et de la mer.

Terrier, C. (2006). 'Flux et afflux de touristes : les instruments de mesure, la géomathématique des flux', *Flux*, 65(3), 47-62.

Vanhoof, M., Hendrickx, L., Puussaar, A., Verstraeten, G., Ploetz, T., & Smoreda, Z. (2017). Exploring the use of mobile phone data for domestic tourism trip analysis. *Netcom. Réseaux, Comunication et Territoires*, 31-3/4, 335-372. https://doi.org/10.4000/netcom.2742

Venturini, T., & Munk, A. K. (2021). *Controversy mapping : A field guide*. Polity Press.

Volo, S. (2020). 'Tourism statistics, indicators and big data: a perspective article', *Tourism Review,* 75(1), 304-309.

Winiecki, D. J. (2008). An ethnostatistical analysis of performance measurement. *Performance Improvement Quarterly*, 20(3-4), 185-209.