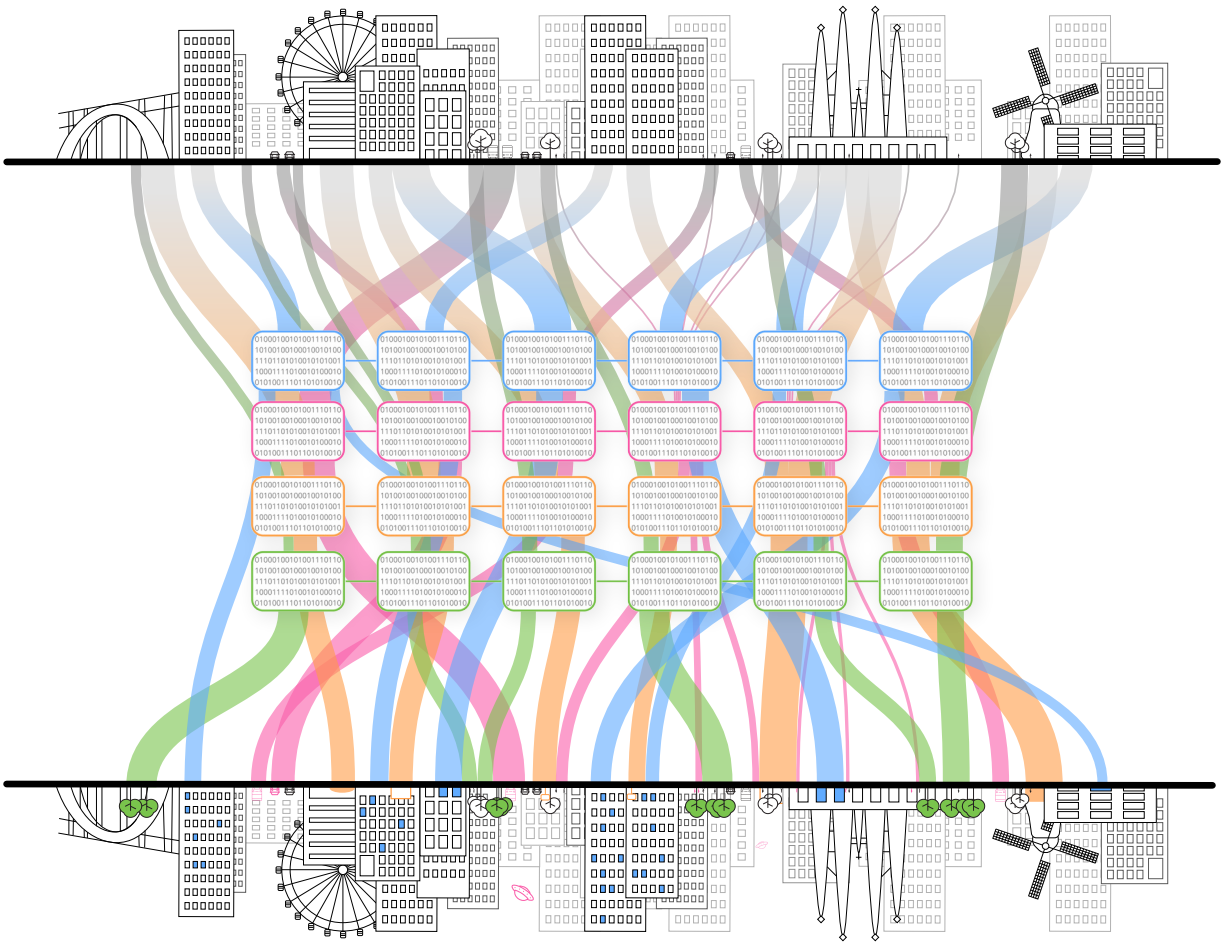
GOVERNING THE URBAN DATA COMMONS

Framework for Methods of an Equitable Governance of Urban Data as a Common Good



Master Thesis at the Chair of Urban Development
TUM School of Engineering and Design
Technical University of Munich
Summer Semester 2022

Caspar Florens Kleiner

Expressions of Thanks

The present thesis has been a remarkable chance for me to explore the topic digital rights and data justice, topics that has been important for me for very long. I am more than thankful for the opportunity to connect it to my field of studies discussing how questions of digital rights affect the built environment and our role as planners in urban contexts.
This would not have been possible without the guidance, support, and critical feedback from my supervisors Prof. Dr. Alain Thierstein and Nick Förster, as well as Mathias Heidinger and Markus Weinig. To whom I would like express my gratitude. Further, I would like to thank Prof. Dr. Felix Creutzig and Prof. Dr. Caroline Nevejan for their critical feedback and helpful comments on the framework and my thesis. It goes without saying that the conduction of the study would not have been possible without the openness and availability of my interview partners in Amsterdam, Barcelona, Porto, and Vienna. I would like to thank you and all the other people I met alongside my journey to your cities for the thought-provoking discussions and the time invested. Likewise, I would like to thank the stakeholders I interviewed in Berlin and Twente for their feedback and stimulating contributions.
I thank my parents Pamela and Christoph for their everlasting support and trust in me throughout my studies, and my brothers Cornelius and Constantin for their enriching discussions and encouragements. Constantin, your feedback and advice have been crucial to me during the past months. I would like to thank you, Daniela, for your love, support, and for always being there. Finally, I would like to thank my friends Anna, Charlotte, Fabian, and Hanna for your help and critical feedback on my thesis.

To you all, thank you very much for your critical feedback, words of support, and your guidance. Pursuing this thesis was a lot of fun and a more than valuable teaching for me!

List of Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface; facilitates interoperability between programs
CC4DR	Cities Coalition for Digital Rights; EU and UN backed coalition for promotion of Digital Rights
CPR	Common Pool Resource; a shared resource that is subject to rules and regulations set by the community consuming it
GDPR	General Data Protection Regulation; EU regulation on data protection and privacy

*Visualisations with an asterix are available as an interactive visualisation by clicking on the caption or by following the link in the table of tables and figures.

1. Introduction

1.1 Abstract

Cities are increasingly relying on data-driven decision-making in urban development issues, like mobility or regeneration projects. Data used in such applications derives from diverse sources with varying methods of gathering, dissemination, and availabilities, for example, through city-owned sensor networks, citizen sensing projects or through purchasing from external data sources. Often, cities cooperate with private businesses that gather, manage, and capitalise on this data. These businesses have control over the data and not only use it for the optimisation of urban development but also as a resource to generate economic value through the data and to influence the behaviour of people in cities. This means an enclosure and potential misuse of the resource data; it becomes inaccessible to the citizens that generated the data in the first place. Control over this data and how it is being generated leads to control over how the cities evolve. Thus, questions of data ownership and stewardship arise. The data subjects, or citizens that move around in a city, and the city governments risk that the urban development decisions are taken out of a political discussion into the control of private businesses and their interests.

As a counter-narrative to this trend often called surveillance capitalism and the exploitation of the good data, the notion of data as a commons is arising in multiple cities in Europe. As such the circumstances of data production, the distribution of the resource, and the uses of the resource need to be controlled and maintained through a political discussion. For this governance of the good urban data, specific methods and measures are necessary to maintain the good as a common pool resource. However, there appears to be no overview of what types of methods exist pursuing this goal.

Therefore, this study examines what kind of methods for an equitable governance of spatial behavioural data in European cities exist and proposes a framework for classifying these methods. For this, the study consists of a scoping review that aims at understanding the scope of the topic *data as a commons in cities*, as well as selected interviews among city representatives and public interest groups or scholars in four European cities pursuing an equitable data governance.

1.2 Contextualisation of the Topic

Mobility, transport, and the ways people move through cities are a fundamental part of the functioning of cities. They shape the face of public spaces, prompt infrastructural demands, and induce services and buildings that form the urban fabric. The different types of mobility, their effects and struggles have always been measured, analysed, and optimised. Be it through traffic counting, mapping of problems or through user evaluation. With the emergence of smart city applications and ubiquitous presence of sensors, such processes are promised to be optimised more efficiently and effectively through means of data gathering and evaluation. This *datafication* of decision making processes has been assigned to help raise the efficiency as well as accountability of urban development decisions in cities (Kitchin 2014).

However, as cities start to collect data in numerous ways, questions of data ownership and stewardship arise. As many of these cities cooperate with private businesses for the implementation of such systems (Hack 2019), or rely on external datasets they buy from private businesses, many of these datasets remain inaccessible to the general public, they are enclosed. Whilst the data is used to influence the behaviour of people in the city by generating new amenities or optimising specific modes of mobility, the decisions about how cities develop become inaccessible to the citizens who generated the data for these decisions in the first place.

The overarching question of who is in charge of how a city evolves becomes relevant; is it the city government through top-down managerial methods, a bottom-up political discussion among its citizens, or is it in the control of private businesses that automate processes and gather data. Businesses might pursue diverging goals from the public interest with the gathering and supply of the data; e.g. preference of one mode of transport, maximisation of commercials being shown to people, or the maximisation of profit through data being gathered. For them, data has become an important economic resource that has to be used to create economic value. Through the analysis of people's behaviour businesses can compile information on who moves how, when, and where in space. This information is valuable for advertising, product development and profiling also in the non-physical space. This represents a potential misuse of data supposed to be gathered for the optimisation of urban processes like mobility and the gathering of data itself. Zuboff (2019) has described this development as "Surveillance Capitalism" and critiques this imperative of regarding data as a resource that must be exploited to create maximum economic return.

As a counter-narrative to this exploitation and extractivism of the good urban data, i.e. data generated with a geophysical relationship to urban environments, the notion of data as a commons has arisen. As data in urban environments is generated by the users, that means people who move around the city, henceforth the data subjects, and the use of this generated data have a direct impact on the way the city works and how the data subjects act in the city, the notion of the commons is used to create an antithesis to data gathering as a means of control (de Lange 2019). Niaros (2016) argues that a commons-approach to smart city projects could help democratising datafication and lead to higher citizen participation. On these backgrounds, there are two problems with an enclosure of the resource urban data; the automation of decisions without a political discussion and the exploitation of the good for economic interests. The notion data as a commons is discussed in the academic literature using diverse terms like "commons-based peer production" (Benkler and Nissenbaum

2006), “knowledge commons” (Hess and Ostrom 2007), or “data as a commons” (Cuff, Hansen and Kang 2008).

Thus, it is necessary to discuss the term *commons*, how a good can be considered a common good and what a potential over- and misuse of common goods could be. The term commons describes common pool resources (CPR) that are managed collectively, meaning collective decision of access to, consumption of, and rules surrounding the resource. Hardin (1968) argued that the unrestricted and uncontrolled use of a resource through many stakeholders inevitably results in an overuse of that good as the individual does not have an interest in limiting their individual use. The only way for Hardin to avert this *tragedy of the commons* is to either privatise the good or through governmental regulations. However, later Ostrom (1990) showed that in practice different modes of governance of CPRs are observable beyond the extremes of privatisation and governmental regulation. These include not only governmental control but also societal and informal methods, specifically for a community self-regulated governance of the resources. For this, she formulated eight principles for governing CPRs, that ensure that all people affected can participate in the defining of the rules (Ostrom 1990: 90). Partial private or public enclosures of the resource, i.e. private ownership of a part of the resource, do not represent a contradiction to the status of a common good of the overall resource, as long as these enclosures are subject to rules, limitations and control of collective governance. Likewise, a good being a CPR does also not mean that everyone has free and unlimited access to the complete resource.

Further, CPRs are split into natural resources like fish, water, or the atmosphere and socio-cultural resources, i.e. human-made resources like knowledge or the city itself as a place of proximity and exchanges. Helfrich (2014) argues that whether a good is a common good is only partially due to its nature but more importantly due to the collective actions that maintain it as such. Thus, any good can become a common good if it is governed accordingly.

In neo-classical understanding, goods are classified in four categories depending on their rivalry/subtractibility and excludability, i.e. whether the use of the resource prevents another use (rivalry) and whether access to the resource can be limited (excludability). In this logic, common pool resources are generally non-excludable and rivalrous. However, for the case of data as a commons Nikander et al. (2020) argue that as data does not diminish, but rather thrives through its use, this understanding should be adopted by a third “anti-rival” category. Herein, data could be considered either a network or a symbiotic good, depending on its governance.

	Rival	Non-Rival	Anti-Rival
Excludable	Private Goods	Club Goods	Network Goods
Non-Excludable	CPRs	Public Goods	Symbiotic Goods

Table 1: Expansion of the Good Classification Model as proposed by (Nikander et al. 2020: 2)

For the case of data as a commons, the question remains whether data is a natural resource that is a CPR a priori, like water or the atmosphere, or whether it is a socio-cultural resource we collectively make a common good through our actions.

D’Ignazio and Klein (2020) argue that datasets are not neutral, but the way in which data is gathered reflects the meaningfulness it has for specific application areas. Pre-existing conceptions and biases influence the outcome of the dataset generated. For example, a traffic counting that only reflects motorised individual mobility and public transport but does not account for bike mobility would thus result in a discrimination of bike traffic if used in an automated decision-making process in general

traffic. Thus, control over data does not only mean enabling decision-making, but also to control these decisions and the meaningfulness of the data that underlies these decisions. Therefore, the question of how data used for urban development decisions is governed and made accessible among the different stakeholders becomes relevant.

1.3 Research Topic and Research Question

Elinor Ostrom's approach and understanding of commons has already been applied to the context of data and specifically urban data; the eight Design Principles for governing commons (Ostrom 1990: 90) have been adapted to mobility data by van Loon and Snijders (2021) or by Chyi and Panfil (2020) for data governance in general. However, these remain very conceptual and do not articulate specific methods that maintain data as a common good. While Ostrom also described exemplary methods of governance of commons in practice, there appears to be no such assessment of urban data as a commons. The evaluation of specific methods appears to be more eclectic with some scholars discussing the role of privacy protection in smart city projects (Kitchin 2016) and others focussing on the availability of data and open data platforms (Walravens, Breuer and Ballon 2014). Data as a commons is discussed as a means of accountability of sustainable urban development by Creutzig (2021) or Labaeye (2019b), or as a way to create citizen engagement (de Hoop et al. 2022).
The state of the academic discussion shows a diverse background of different disciplines. In his dissertation, Labaeye (2019a: 53) points out the need for more research in the field of the institutionalisation of commoning practices. He stresses the special role of municipal governments that could lead to an up-scaling of commoning practices. This research gap is where this study tries to contribute with the formulation of a framework for the typology of different methods for the commoning of urban data.
With the notion of urban data as a commons on the rise, the question arises what methods exist to maintain this good as a CPR. It appears like while the idea of data as a commons is actively discussed, it remains unclear what specific governance methods this idea translates into. It seems like the process of discussion around the conditions of data gathering, and the power relations between cities, citizens and private businesses are not yet set and the balancing of specific rules governing the good urban data is not completed. While there are specific laws that address data regulation, like freedom of information or archiving laws, it seems like there is no general overview of what types of methods exist, being proposed, or already in use, for the governance of urban data as a commons. For this reason, the study focusses on this question in the context of spatial behaviour data, i.e. one type of urban data referring to geo-spatial data being gathered in urban environments reflecting human behaviour in space, in European cities and proposes a framework for categorising different methods of governance of urban data as a commons.
Thus the research question of this study is:

> What types of methods exist for an equitable governance of spatial behavioural data as a common good in European cities?

Hypotheses:

In the discussion of urban data as a CPR, the conditions of its production, rules of its use, and the power relations between people, governments and private businesses are not yet well defined.

For considering data a CPR, methods of governance are required that empower data subjects to access data and the public discussion around its production, use and regulation.

The borders between completely enclosed data, (partially) open data, and data as a CPR are vague and fluent.

The different stakeholders regarding this question are the citizens or data subjects, private businesses that gather data and the city governments. The study focusses on spatial behaviour data as this closely links to personal data rights. Examples for this kind of data can be data on movement of pedestrians in the city, or by means of transport like bicycles, cars, or public transport. Governance does not only refer to the legal framework that exists through laws and regulations, but furthermore also to the (unwritten) rules and procedures that maintain the good, the conditions and structures of power between the different stakeholders in society (Bevir 2012). For urban data this governance does not only include governing the data being controlled by the city government itself, but also the data being held by private businesses and individual data subjects (von Grafenstein, Wernick and Olk 2019). Equity can be defined as not only distributional but also as procedural fairness (Leventhal 1980). That means not only the accessibility to the resource and its lawful collection but also an accessibility to the discussion about its circumstances and governance. In the case of data, this can be translated as three dimensions of governance that need to be subject to a societal discussion:

1. The Conditions of Data Production
2. The Distribution of the Resource
3. The Use of the Resource

The governance of data is thus equitable, when it permits societal decisions of these dimensions.
Methods allowing this equitable governance can be diverse from rules that define what data can be gathered to incentives for data sharing or technical education about data gathering. A first overview of the literature and research about the topics suggests that there could be four different types of methods relevant:

(1) Legal Methods that define the basic rules of data gathering and the responsibilities of stakeholders, e.g. the General Data Protection Regulation (GDPR) that limits the gathering of personal data.

(2) Input-Side Methods that inform and communicate about data gathering in urban environments, e.g. street signs or information brochures.

(3) Output-Side Methods that address the distribution and accessibility to the data, e.g. open data portals or hackathons.

(4) Design Principles that set basic procedures for data governance, e.g. privacy by design or data minimisation.

The aim of the study is to identify methods of an equitable governance of urban data and to develop a conceptual framework for classifying these methods. The use of this framework is both to help

cities to identify methods they have not implemented yet, and to systematise the discourse of different methods of an equitable governance of data in cities.

2. Methodology

As the topic of equitable data governance exists both as a theoretical topic in the academic discourse as well as a concrete responsibility cities are facing, the exploration of the research questions requires an engagement of both the academic literature and practical implementations. Therefore, the study is split into two methodological sections; a scoping review of the academic literature and expert interviews with suitable partners in European cities.
Following the research question and the diverse background of the academic discussion, the scoping review aims at delineating the extent of the research question in the academic fields. For this the scoping review intents at answering the following questions:

- What disciplines are active in the field and what topics are relevant?
- How are these topics discussed?
- How can data be understood and managed as a common good?

Further, to understand what types of methods exist, it is first necessary to learn what methods exist. Thus, the research question provokes the following questions that the interviews with stakeholders in the field target to answer:

- What methods of governance of the good data are used?
- Is data understood as a CPR in European cities?
- What difficulties and limitations in the governance of data as a common good are relevant in practical discourse?

2.1 Scoping Review

For delineating the extent of the research topic, fitting literature is identified via specific search terms in common scientific literature databases. Given specific exclusion and inclusion criteria, the identified literature is narrowed down to an eligible selection. Subsequently, a quantitative bibliometric analysis as well as a qualitative discussion of the emerging concepts are conducted.

2.1.1 Identification of Literature

The collection of the literature is identified through queries in the scientific literature databases Scopus and Web of Science. Additionally, a search via Google Scholar is conducted. Following the recency of large-scale data gathering in smart city projects, the queries seek to identify literature

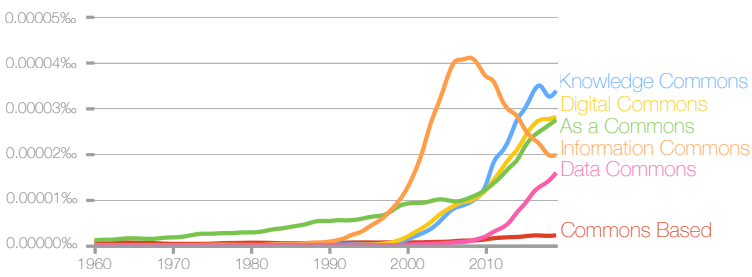


Figure 1: Frequency of the Terms in English Literature (1960-2022)*, plotted in Google Books Ngram Viewer (Google 2022)

published between the years 2012 to 2022 to identify a maximum amount of literature relevant. Using the terms for data as a commons identified prior, these and similar terms are assessed for their sensitivity, i.e. whether they return relevant and not too broad results. The terms are then gathered with the association to “urban“ and “data governance”. Using the advanced search fields, the following terms are searched in the databases:

Scopus: *n*=29[1]

"Data Commons" OR "Knowledge Commons" OR "Information Commons" OR "Digital Commons" OR "Commons based peer production" AND urban "Data Governance"

Web of Science: *n*=137[1]

TI=(Commons AND (Data OR Knowledge OR Information OR digital OR based peer production) AND Urban AND Data Governance) OR TS=(Commons AND (Data OR Knowledge OR Information OR digital OR based peer production) AND Urban AND Data Governance) OR AB=(Commons AND (Data OR Knowledge OR Information OR digital OR based peer production) AND Urban AND Data Governance)

Google Scholar: *n*=516[1]

"Data Commons" OR "Knowledge Commons" OR "Information Commons" OR "Digital Commons" OR "Commons based peer production" AND Urban "Data Governance“

The identified references are searched for doubles and then evaluated according to the following criteria.

Exclusion criteria:

Language: References that are not English, French, or German are excluded.

Type: References that are not books, book sections, journal articles, and conference proceedings are excluded.

Title: References that are not referring to data governance, urban data or data as a commons are excluded.

Availability: References that are not openly accessible, or accessible to the author through libraries are excluded.

Abstract: References that do not cover the topic data as a commons, that do not discuss themes relevant for urban data, or geospatial behavioural data are excluded.

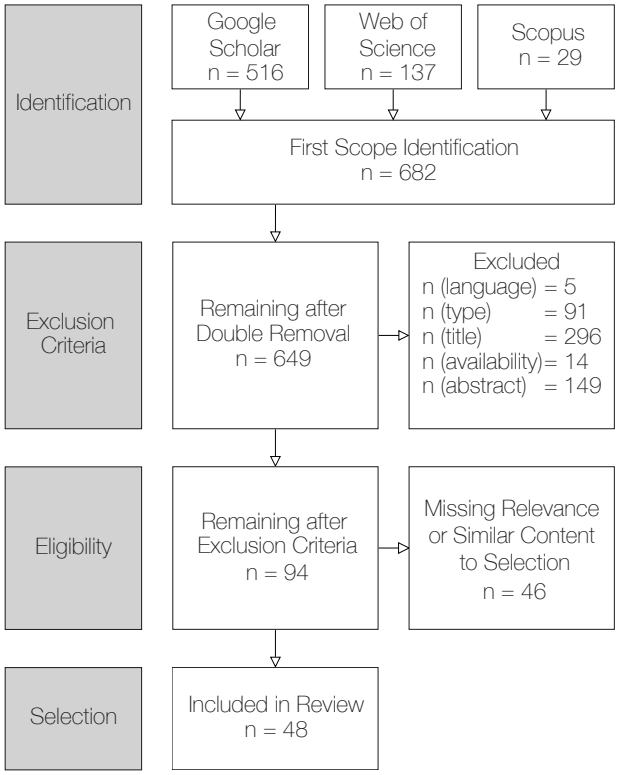


Figure 2: The Study Selection is plotted according to PRISM-Flow (Tricco et al. 2018)

Inclusion criteria:

Relevance: Only references relevant for the research question are included. From multiple references covering the same concerns only the most relevant ones are discussed in the review.

[1] Search conducted in May 2022

2.1.2 Bibliometric Analysis and Review

The identified literature is categorised by the different fields of studies as well as the different topics they cover. The fields of studies are identified by the journal in which the academic articles are published, or in the case of books and journals that bridge multiple fields through the majority of the fields of the authors. The topics covered are determined by the keywords given through the authors as well as an examination of the reference abstracts. The field of studies is assigned exclusively for each reference and the topics covered inclusively, meaning that each reference is assigned to only one field of studies, whereas the topics can be multiple topics for each reference. For this part of the review all the references eligible are included. In a second step, the frequency of the different fields of studies and the topics discussed are evaluated. Their co-occurrences are visualised, analysing which fields cover which topics. Further, the distribution over the temporal course is analysed as well as the interrelation of the topics.
Subsequently, the topics identified in the bibliometric analysis are examined and discussed individually. Here, only the most relevant references are discussed. The topics are examined according to their content-related arguments and how the different concerns and topics voiced in the references relate to one another.

2.2 Expert Interviews

The expert interviews with stakeholders with both city representatives, i.e. people working in a city government body relating to the topic of the study, as well as public representatives, i.e. both scholars focussing on the topic as well as representatives of public interest groups, are shedding light on two areas. On the one hand, they generate insight in what methods cities are already implementing or will implement in the future and how these methods are perceived from both sides. On the other hand, they will give information about whether data is considered a common good or not, what the main struggles and limitations are and what risks are connected to the topic.

2.2.1 Identification of Cities

Cities that implement smart city projects and make their data partially available via open data portals are innumerable. Cities that go further and discuss the role of digital rights, data ownership and accessibility are scarcer. The Cities Coalition for Digital Rights (CC4DR)[2] founded in 2018 by the cities of Amsterdam, Barcelona, and New York together with the United Nations and the European Union pursues to “work towards legal, ethical and operational frameworks to advance human rights in digital environments” (Cities Coalition for Digital Rights s.a.) and helps therefore to identify fitting cities for interviews. Among the 54 cities having joined the coalition ever since, both the cities of Amsterdam and Barcelona are positioning the concept data as a commons in their digitisation agendas. In Amsterdam it is approved as a “task that is elaborated with partners, knowledge institutions and experts from the city” (Gemeente Amsterdam 2019: 27, translation by the author)

[2] citiesfordigitalrights.org last accessed 12.07.2022

and as a key strategy in Barcelona (Ajuntament Barcelona 2018) that is known for its strong citizen empowerment approach to smart city and data gathering projects (Calzada 2018). The two cities are thus more than relevant case studies. Besides these two cities explicitly emphasizing the concept data as a commons, two more cities are interesting based on the way describing their digitisation projects. This goes for the city of Porto due to the high focus on ethics and privacy but also on participation and accessibility to data (Campolargo and Calçada 2018). It also applies to the city of Vienna because of the focus on transparency and inclusion of all citizens in the digital agenda (Stadt Wien 2019) and the *Data Excellence* approach that sets the rule for data governance in the city (Lutz 2019). The four cities thus give an overview of both cities that explicitly frame their approach as data as a commons as well as cities that strongly focus on civic empowerment. For each city fitting interview partners within the city governments as well as public representatives are identified and requested for an interview.

	Amsterdam	Barcelona	Porto	Vienna
City Representatives	4	1	3	2
Public Representatives	2	2	1	1

Table 2: Number of Interview Partners in Respective Cities

2.2.2 Interview Conduction

The majority of the interviews is conducted in person, with only a few exceptions that had to be held online. The interviews are conducted semi-structured as open interviews following a general outline that is prepared beforehand and is the same for each interview. However, for each interview a preparation for the individual interview partner's background and field reveals specific and relevant questions and topics. The general interview outline is split into two sections. First, a general discussion about data as a commons aims at revealing the interviewee's understanding of data as a commons, and also how data can be considered a common good. The section further explores what limitations, difficulties and threats of data as a commons and digitisation in general the interviewee determines to be relevant. In the second section, the focus shifts towards the specific methods of data governance relevant. Here, the outline follows the three dimensions of equitable data governance; conditions of data production, distribution, and use of the resource. Specific questions to identify potential methods are formulated here. A third area in the interview outline focussing on a temporal dimension and prioritisation of methods emerged as too soon to be asked in most interviews and is thus not respected in the evaluation. The purpose of the interview outline is not to give a strict guideline that needs to be followed, but rather a checklist on topics than can be discussed if deemed relevant. The full outline for the open interviews can be found in the annex of the study (9.2).

2.2.3 Qualitative Analysis of the Interviews

Besides the assessment of the interviews for methods mentioned, the interviews are also analysed qualitatively following the qualitative content analysis by Mayring (2015). For this, the interviews are recorded and fully transcribed afterwards. The spoken contents are marginally adjusted to a more readable and natural style, i.e. irrelevant and unfinished parts are removed and some inaccuracies are corrected. After the transcription of the interviews, it becomes clear that besides the types of methods another criterion for classification seems to be relevant; that is the aims of these methods. In regard to the relation between private businesses and public administration to the citizens, four themes can be isolated:
Data Protection, Collecting of Consent, Creating of Awareness, and Enabling of Interaction.
Subsequently, besides the general discussion of data as a commons and methods of data governance, the interviews are also analysed for these four categories. For this, the transcribed interviews are coded using the qualitative data analysis software ATLAS.ti. Here, the statements in the interview are attached to codes that condense the contents of the statements. These codes are developed inductive-deductively, i.e. partially prepared beforehand and partially through the process of reading and assigning codes to the interviews. While the interview outline already includes a section of the general discussion and a section for specific methods, the analysis is conducted independently from these sections, analysing what the interview partners mentioned, independent from the position during the interview.
After all interviews are analysed accordingly, the codes are arranged and joined where necessary. The coding outcome is presented and the topics emerging are evaluated accordingly. For this, the generated interview data is visualised and assessed. The interviews are first evaluated generally pointing out to general topics and differences in the interviews between city representatives and public representatives and then evaluated for each city individually.

3. Scoping Review

3.1 Bibliometric Analysis

The identified literature eligible for the bibliometric analysis consists of 94 publications stemming from diverse fields of studies and focussing on a multitude of different topics. Figure 3 shows the rate of each field of study, the topics as well as what fields covered which topics.

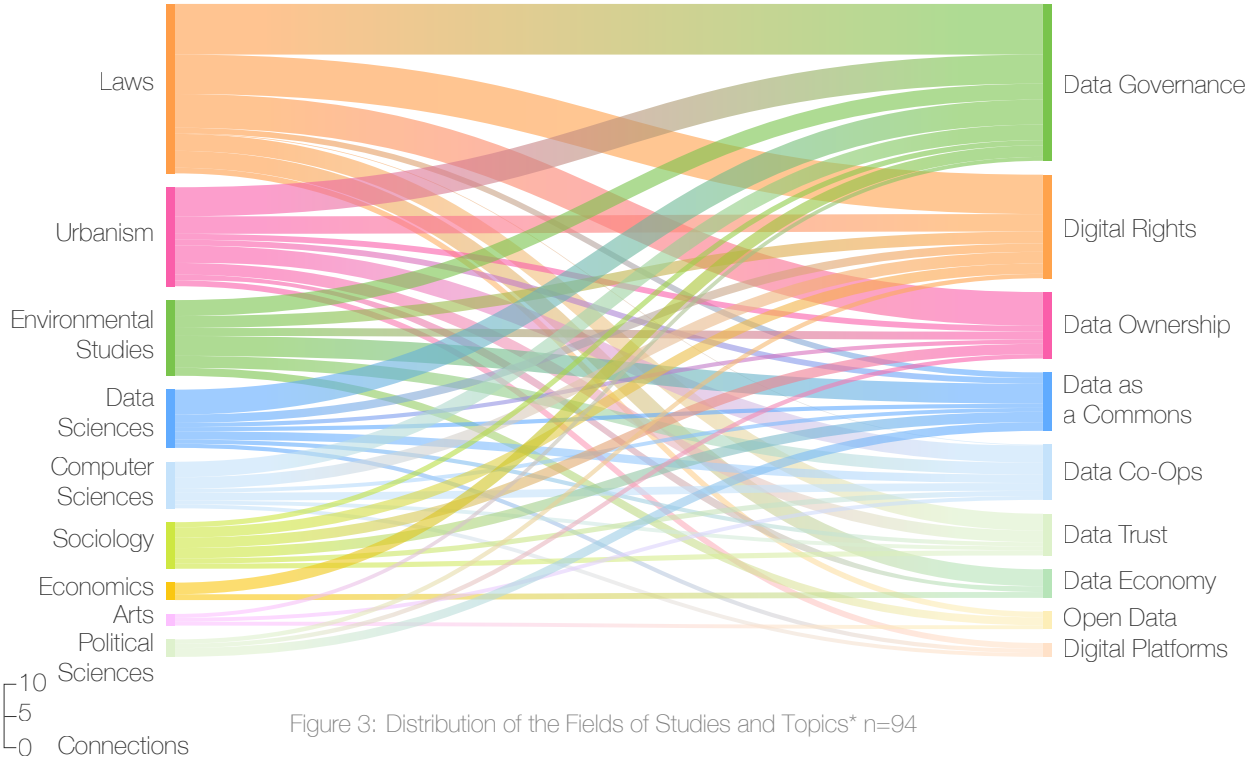


Figure 3: Distribution of the Fields of Studies and Topics* n=94

Predominantly, Law scholars seem to shape the discussion with 29 publications, with a distance to the two next fields Urbanism with 17 and Environmental Studies with 13 publications. Data Sciences, Computer Sciences and Sociology appear to play an equally big role between eight to ten publications each. Surprisingly, Political and Governmental Sciences appear to play a subordinate role, especially as these three publications are covering the topics Data as a Commons, Data Ownership, and Digital Rights rather than Data Governance. There are references refering to multiple topics, thus doubles occur. The topics appear equally diverse as the disciplines. Here, it appears Data Governance in general is discussed most often, followed by Digital Rights and concerns of Data Ownership. With Data Cooperatives, Data Trusts and Digital Platforms specific methods already emerge as topics. References discussing the topic Data as a Commons appear less often, however stem mainly from the field Environmental Studies. Data Governance, Digital Rights, and Data Ownership are the main topics discussed by Law scholars, whereas Urbanism scholars seem to focus on Data Cooperatives than on concerns of Data Ownership. Besides Urbanism, Data Cooperatives are mainly addressed by Environmental and Data Sciences. Data Sciences scholars conversely focus more on the topic Data Governance than scholars in the field of Environmental studies. Data Economy is primarily addressed by Law and Economics scholars and interestingly the topic Data Trust both by Law and Urbanism scholars.

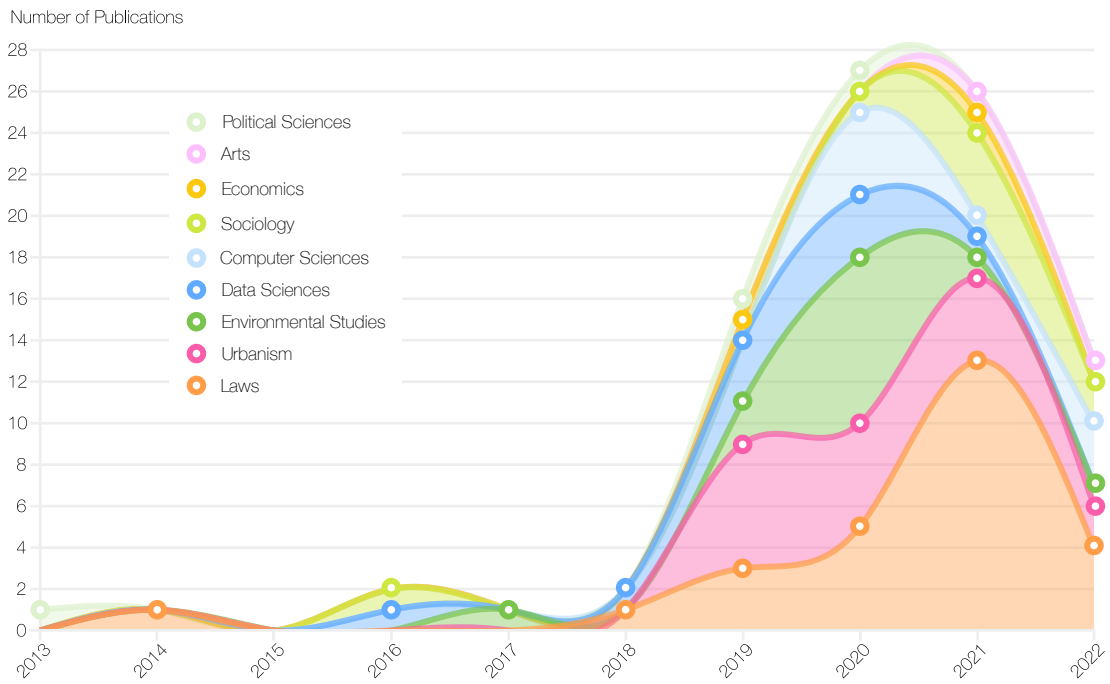


Figure 4: Field of Studies by Year*, stacked, only points visible if publications in respective field

If looked at the fields of studies from a different perspective by contrast, figure 4 shows that the relevance of the different fields has changed over the years. With only occasional publications in different fields prior, the discourse of governance of urban data as a commons appears to start in 2019 accelerated by six publications in the field of Urbanism. In the same year, among the identified literature there were three publications from the field of Laws as well as two from Environmental Studies and Data Sciences each. So far, in 2020 the most references have been published. During that year there were more publications in the field of Environmental Sciences (=8) than in the field of Urbanism and Laws (both 5). However, in the discipline Computer Sciences the topic gained more traction as opposed to Data Sciences in 2020. Only in 2021 is the distribution clearly shifted to the field of Laws with 13 publications. The fields Urbanism and Sociology with four publications each play a subordinated role. For the year 2022, this trend appears to be continuing with four publications in the field of Laws, three in Computer Sciences and two each in Urbanism and Sociology. However, as the literature identification was conducted in May 2022 the number of publications is not yet representable for this year.

Plotted in the same way as figure 4, the distribution of the different topics covered in the identified literature has stayed largely the same each year as opposed to the shifting frequency of fields of studies. Here, the different topics are surprisingly equally often addressed each year as they are addressed in total. Only the topic data cooperatives is covered relatively more often in 2020. Figure 5 shows the overlaps of topics in the identified literature. Mainly the topics data-cooperatives and data governance and data ownership with data as a commons and digital rights occur together.

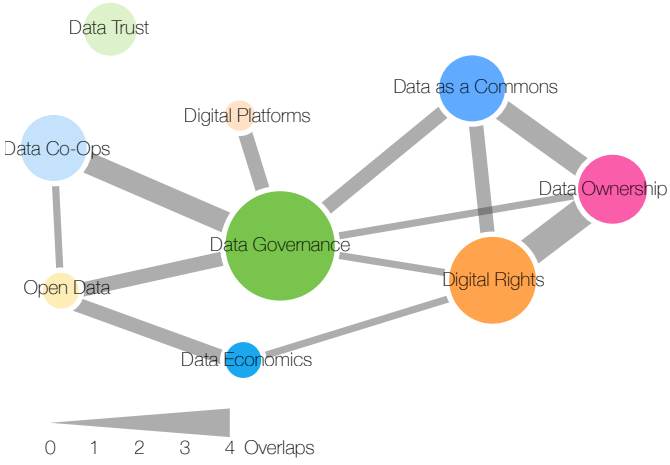


Figure 5: Overlaps in Topics*

3.2 Review

The aim of the review of the identified literature is to understand the extent of topics and how they are discussed. For this, the main concepts and different approaches to concerns are discussed and juxtaposed. As opposed to the bibliometric analysis, only the 48 most important and relevant references are discussed. These references cover all topics and concerns identified in the literature in the most comprehensive manner.

3.2.1 Data Governance

Data governance as described earlier refers to both the legal and institutional but also unwritten rules and definitions of power structures between stakeholders surrounding data.
Micheli et al. (2020) identify four distinct models of data governance countering the predominant model of corporate extractivism and platform data governance; data sharing pools, data cooperatives, public data trusts as well as personal data sovereignty. These governance models are examined and compared using five criteria; the stakeholders involved, governance goals, value from the data, governance mechanisms, as well as the reciprocities, i.e. power relations between the stakeholders. On this background, data sharing pools are described as infrastructures between mainly business entities aiming at improving the value of the data and therefore criticised for excluding public interests. Data cooperatives on the other hand are discussed as institutions where individual data subjects can decide what their data is used for, thus improving the self-determination of the data subjects. Public data trusts are examined as places where multiple public bodies like government offices save and share the data they have about data subjects centrally to improve the useability of the data. Personal data sovereignty is described as intermediate services where data subjects can individually govern their data to be shared with businesses, however criticised for not scaling and focussing on creating private profits. As the four models are not used broadly, the study points to the main problem that there needs to be proper regulation to incentivise these models of governance.
Following this identification of governance models, Zygmuntowski, Zoboli and Nemitz (2021) classify the different models based on their governance goals as well as governance mechanisms.

	Mechanisms	
Goals	Individual: rights-based	Institutional: trust-based
Private: growth-driven	Personal Data Sovereignity	Data Sharing Pools
Public: welfare-driven	Data Cooperatives	Public Data Trusts

Table 3: Goals and Mechanisms of Governance as proposed by (Zygmuntowski, Zoboli and Nemitz 2021: 8)

By classifying the models, they criticise the commodification, and thus continued imperative of extractivism, of data, as well as the neglect of public interest in the governance models. To address the public interests as well as protection of individual rights, the authors propose the governance of data as a CPR following Ostrom's Design Principles (Ostrom 1990: 90) in connection to the issues of privacy, value, control, access and exclusion (Zygmuntowski, Zoboli and Nemitz 2021: 19-20).
However, Prainsack (2019) argues that the governance of data as a CPR in the same way as traditional CPRs does not sufficiently address the problem of digital exclusion and criticises that already established stakeholders in the data economy profit the most. Focussing on health data as

a commons, the author suggests an adaptation of Ostrom's Design Principles for addressing exclusion by focussing on process, and effects of such exclusion.
This aligns with Madison (2020) who argues that before defining specific rules of governance, the institutional design, i.e. the overall attitude and value system defining data governance, needs to be defined.
For the specific case of spatial behaviour data, Smichowski (2019) concludes that not only one model of governance alone can address all the questions related to accessibility and rights. The author identifies four general models for an equitable governance; (1) Crowdsourced Data Commons, in which many individuals gather data and aggregate it together. (2) The Collective Bargaining over Rights, i.e. collective action towards businesses gathering data. (3) Data Requisition, incentivising data sharing through services, payment, or regulation. (4) Data Pooling, i.e. combining datasets from numerous sources for mutual value amplification. Smichowski argues that the true potential for these governance models for each specific application area is reached if the interactions between the models are addressed and implemented purposefully.
Arriving thus in the specific context of data governance in smart city projects, Le Breton et al. (2022) draw experiences from a case study on collective data governance in Rennes, France. In the balancing of data governance among the stakeholders, the model developed from an open data platform with unstructured data availability, via a centrally organised service for data ("Service Public Métropolitain de la Donnée"/SPMD; metropolitain public service for data (Le Breton et al. 2022: 67)) to a governmentally supervised general interface for data sharing. However, the authors identify main shortcomings in this case study in lacking inclusion of citizens and insufficient addressing of data sovereignty.
To address such challenges diverse solution approaches are discussed in other references focussing on urban data governance; Johnson et al. (2022) identify main considerations generally addressing questions of accountability, transparency and supervision of smart city projects. They subsequently formulate three key elements that need to be respected for successful development: clear definition of the goals, involvement of stakeholders as well as maintaining of flexibility and iterative development. To allow such a stakeholder involvement, Foth et al. (2021) explore the concept of a *DateCare Space*, a physical space where citizens can be informed and get educated about data gathering in the city. They argue that such a space could address four scenarios; raising awareness over data ethics, raising data literacy, engaging participatory assessment of data, and envisioning city data futures (cf. Foth et al. 2021: 330). Lupi (2019) concentrates on a *City Data Plan*, like a city manifesto that delineates the main aspirations, values, and responsibilities a city government is giving itself. The author argues that such a City Data Plan can build accountability and transparency for citizens but also help internally to make decision-making more proficient.
Choenni et al. (2022) on the other hand, argue that whereas today data gathering projects usually focus on specific applications scenarios like electric mobility, for the overall governance specific rules should already be set today. They argue for the *Findable, Accessible, Interoperable and Reusable (FAIR) principles* as a framework to ensure a responsible handling of data in smart city applications.
More concretely, Benfeldt, Persson and Madsen (2020) identify six distinct problems facing practical implementation of data governance through interviews with representatives in 13 Danish municipalities. These consist of perception of value, overview of data existing, fostering capabilities, existence of local practice, enabling collaboration and the political ambience. Further, they map the

interactions of these difficulties on a triangular graph introducing specific approaches for addressing these difficulties (cf. Benfeldt, Persson and Madsen 2020: 308).

3.2.2 Digital Rights

Among the publications in this section, the majority refer to the right to privacy in data gathering contexts. However, Iaione (2019) as well as Fisher and Streinz (2021) argue while existing legal discourses focus primarily on regulation of transfer, protection, sharing, exploitation of data (Fisher and Streinz 2021) as well as privacy, discrimination and security (Iaione 2019), rights of control over, access to data and also participation are underrepresented. Fisher and Streinz conclude that laws cannot fully encompass question of societal "digital destinies" (Fisher and Streinz 2021: 955), whereas Iaione suggests a fivefold governance model comprising of policy, institutional, legal, financial, but also urban design dimensions that target the rights of accessibility and participation to data gathered.

Similarly, however focussing on the right privacy, Benjamin (2020) addresses the limitation of *privacy protectionism*, i.e. the approach of considering privacy as something that has to be protected through Legal Methods, and paints privacy as a performative act. Drawing from Judith Butlers' Performative Acts and Gender Constitution (1988) the approach of this "performing privacy (…) is to critique existing structures of data collection and exploitation with new utterances of identity and consent in socio-relational networks built on choice, agency, and respect" Benjamin (2020: 21), i.e. rendering privacy an individual decision rather than a regulated imposition.

Gaining insights from citizen questionnaires on data privacy practices in smart city projects in Long Beach, USA, Gwen Shaffer argues that city governments "must put equal effort into fostering trust, practicing transparency, and engaging the public" (Shaffer 2021: 251) as into the development of smart city projects. In the study, the author identifies three distinct smart city technologies creating mistrust of citizens into their city governments specifically over concerns what the data is used for and welcomes the city's participation in the CC4DR. To address these questions of mistrust Pierri and Wiltshire (2021) draw experience from the Citizen Voices for Digital Rights that is part of CC4DR. In the study, they identified three main dimensions that are relevant for building trust: the importance of the lived experience, i.e. the added value through smart city projects, the role of digital literacy, and the role of the municipality in offering democratic ways of civic engagement.

Goodman (2020) on the other hand groups overarching expressions of mistrust into three concrete concerns: Privatisation, Platformisation, and Domination. Privatisation referring to the control of data and urban development in the hand of private businesses. Platformisation as the city as a service, i.e. the commodification of urban environments and qualities, and domination as a way of control of people through data gathering and automatic decision-making. The author argues that for a democratic governance, rights of access and control to data need to be guaranteed to ensure accountability and information about data gathering and automatic decision-making.

Drawing from a questionnaire directed at representatives of 13 cities participating in the CC4DR (including Long Beach), Calzada, Pérez-Batlle and Batlle-Montserrat (2021) postulate the five digital rights most relevant for these cities: (1) equal and universal access to the Internet, (2) privacy, Data Protection, and security, (3) transparency, accountability, and nondiscrimination in data, content, and algorithms, (4) participatory democracy, diversity, and inclusion, (5) open and ethical digital

service standards (cf. Calzada, Pérez-Batlle and Batlle-Montserrat 2021: 12). The authors describe that in addressing these rights, the cities are not only implementing rules on their own but also learning from one another. In another publication, Calzada and Almirall (2020) point out this necessity for cities to share data with other cities in order to address the needs of digital rights together as they cannot be addressed by each city individually.

3.2.3 Data Ownership

The references identified under the topic Data Ownership primarily discuss questions of economic justice manifested by the question who owns the data; be it the fair distribution between different data consumers, data gatherers, the community, or the data subject itself.
Taking the example of the American company Uber also operating in India, Singh and Vipra criticise the "current default that 'who collects data owns it'" (Singh and Vipra 2019: 54). This results in data being taken away from the communities in which it was produced. They address the problem of ownership of data as an economic injustice in two dimensions: between businesses that gather data and communities and individuals data is gathered from, as well as geo-politically between countries where these businesses are located and countries where they operate. They therefore argue that data should be owned collectively to be brought to the most uses in the respective communities.
Hummel, Braun and Dabrock (2021) by contrast argue that the concept of ownership is difficult to apply to the resource data. However, they acknowledge the calls for questioning the status quo of data access and distribution but argue that these calls are diverse in their objectives and critiques of the status quo. They argue that the question of ownership or property of data is not actually the concern, but rather a vehicle to voice concerns. The authors thus map four conceptual poles resulting in the call for questioning data ownership and what aims they pursue: (1) "Property-Quasi-Property" with the aim to control data flows and outcomes, i.e. data stewardship. (2) "Marketability-Alienability" with the aim to benefit from the resource and reduce harm, i.e. the freedom of the individual to market their data. (3) "Protection-Participation" with the aim at only disclosing information at the individual's discretion, i.e. protecting individual privacy. (4) "Individual-Collective" with the aim to harmonise individual and common good, i.e. generating overall value (Hummel, Braun and Dabrock 2021: 565). The authors conclude that the calls for ownership might even hinder the objectives as the redistribution of the resource data would not automatically fulfil these objectives.
As a legal scholar, Drexl (2021) argues that the question of economic justice through collective data ownership can be seen through multiple legal lenses. First, it can be considered a topic affected by contract law, meaning as an issue between the data subjects and the businesses or institutions that gather data. On the other hand, it can also be seen as a topic affected by competition law, i.e. as an unfairness between the data gatherers and further data consumers that rely on the data gatherers. Drexl argues however, that a third lens is relevant and thus proposes an additional regime as part of fair-trading law, i.e. a right for data access from both data subjects' perspective to the data they have generated but also from the data consumer's perspective for the value the data could constitute for them. This could for example result in need for technical standards like data formats or an Application Programming Interface (API) in specific sectors as mobility or energy.

Fia (2021) equally recognises the limitation of competition law to address concerns of data access rights. Focussing on what the author refers to as *Raw Non-Personal Data*, they explore the notion data as a commons and differentiate between the structuralist commons approach according to which data access should be granted simply because it is currently enclosed and the functionalist approach focussing on the value the enclosed data could create for the common good. Fia concludes that for a commons-based approach, there needs to be a review of traditional intellectual property protection regulation as well as a new concept of ownership as a set of varying rights instead of comprehensive ownership.
Similarly following this critique of traditional ownership concepts for the case of data, Hicks (2022) explores the concept of collective ownership of data. The author argues that the question of data ownership should be regarded as a question of social interaction rather than as regulatory design problem, i.e. question of empowerment of citizens instead of determining. She formulates four research fields that need to be explored in this regard; focus on non-Western legal concepts of collective ownership, assessment of existing methods and social relations of data governance, focus on communicative process rather than solution of specific problems, and a closer examination of the rhetoric used in the data commons discourse (cf. Hicks 2022: 3; 8-12).

3.2.4 Data as a Commons

The overarching concerns in literature focussing on data as a commons refer to privacy and data protection as well as sustainable development.
Jung Marques et al. (2021) study the relation between commons theory and knowledge-based development focussing on urban data specifically by conducting a literature review in both fields. They conclude that the main overlaps are in value creation as well as in co-production of data, i.e. having a commons-based approach for data that can be used to serve the common good and for data that is generated by a collective. This relates to the functionalist approach to commons that data should be available as it can serve the common good described earlier by Fia (2021).
Such an example for a commons-based approach for serving a common good can be found with Creutzig (2021) who argues that a commons-based approach to data governance of mobility data could help accelerating sustainable development. Creutzig argues that data deriving from shared mobility services like e-scooters should be managed collectively in an integrated data platform. This data can then be used to make decisions for sustainable traffic and make these decisions verifiable towards political goals like sustainability.
Further exploring this idea of accessing data collected by private businesses but could be serving a public good, Delcroix (2017) focusses on how to make personal data and public use of it compatible referring to the French project The Platform of a City. The author sketches four possible paths for the future: two scenarios requiring new legislation (1) mandatory private open data, where data gatherers have to anonymise private data and make it available freely, (2) Enhanced general interest in data, where pseudonymised private data is made accessible to public actors like government entities. As well as two scenarios within existing legislation (3) data reuse platform, in which personal data is aggregated and made freely available, or (4) citizen portability, where the citizens themselves decide to make their data available to public actors. Finally, Delcroix points out to missing specific

methods of governance of data as a commons for bridging the concerns over personal data and privacy.

Wong, Henderson and Ball (2022) study the hindrances of privacy and Data Protection in the development of data as a commons more closely. For this, they interview commons experts and conclude that privacy is primarily considered a question of empowerment of the data subjects. Consequently, Wong, Henderson and Ball propose a checklist according to which agency of data subjects and their Data Protection can be ensured. This checklist primarily foresees co-creation and collaboration with the data subjects and defines four dimensions; identifying of the commons and the data subjects, scoping and information gathering for developing the commons, building the commons, and sustaining the commons (cf. Wong, Henderson and Ball 2022: 23-25).

Park (2021) describes how traditional enclosures of goods occur to address specific societal concerns regarding the resource, e.g. the privatisation of traditional common land to ensure the production of livestock on it for all. The concern driving for an enclosure/privatisation of the good data, Park argues, has been a mechanism to ensure privacy protection. However, the author examines that the concept of private ownership does not avert this tragedy of the resource. Conversely, Park clarifies that an enclosure of the good data itself leads to its depletion rather than an opening of the resources as economic interests of the data holders incentivise not to make it available thus withdrawing the resource from common benefit.

On the other hand, Taylor (2016) questions what this argument for the common good or common benefit refers to in real world instances and explores the hindrances of private businesses to disclose. The author identifies these hindrances primarily because of economic interest of the businesses to scarcen the resource thereby creating an advantage over other stakeholders and creating maximum economic value for the own business. Further, even if the data does not create an advantage for the business, Taylor argues that the benefit and the contextual value of data if shared are not visible to most data holders. The author concludes that for the data to generate a maximum value for the public good, it should be considered a public good, stating however that the "responsible data debate has not been able to overcome the obstacles (…) by producing technical and ethical frameworks within which data can operate as a public good" (Taylor 2016: 11).

3.2.5 Data Co-Operativism

Under the topic Data Co-Operativism both citizen-initiated data co-operatives, city-initiated collaboration with citizens, as well as citizen science are discussed.

Morell, Cigarini and Hidalgo (2021) explore the value of citizen science, i.e. data and knowledge gathering through laypeople, for collective governance of scientific data and knowledge in general. They analyse specific commons-oriented citizen science platforms and identify three dimensions they address relevant for collective control, systematising data co-operatives: governance, i.e. institutionalising ways of community feedback, knowledge and technology policy, i.e. reflection of what is gathered and standardisation of data, as well as effects, i.e. focus on the impacts and responsibilities the gathered data generate. Thus arguing that commons-oriented citizen science co-operatives can ensure ethical and effective data governance.

Van Zoonen (2020) similarly points out to these effects citizen involvement can have and calls for modes of civic participation in urban digitisation projects. The author notes that the main struggles such projects face today are their existence within an institutional void (no regulation etc.), intransparencies regarding the quality of data they generate and use, and also conflicts with personal data protection. Further, van Zoonen argues for the necessity of involving citizens and specifically addresses the notion of data as a CPR for the case of data produced in urban environments.
Leclercq and Rijshouwer (2022) focus on the topic from the opposite perspective, regarding the right of citizens to participate in urban digitisation, and the role of digital platforms. They argue that while citizen involvement can augment citizen engagement, empowerment, and emancipation, however such processes of citizen involvement are still considered as requirements cities have to meet rather than a resource cities can harness for efficiency and building of union among citizens and city governments.
Focussing on a Finnish data activism project, MyData, Lehtiniemi and Ruckenstein (2018) observe that data collaboratives emerge as a response to concerns of data exploitation and inefficiencies. They label two different imaginaries, i.e. a set of values and imageries of how something works, how data collaboratives address these concerns; technological, i.e. as a corrective measure to the concern, and socio-cultural, i.e. as a monitoring of the effect of the concern and the corrective measure. They argue that diverging concerns and concepts of the societal well-being exist from different stakeholders, like governmental, engineers or sociologists, and thus conclude that "we need to keep asking what productive 'critical engagement' means in the context of data activism and developing (of) data infrastructures" (Lehtiniemi and Ruckenstein 2018: 10)
Calzada (2020) has a broader look at platform and data co-operatives and focusses on co-operatives accelerated through the COVID-19 pandemic in three European cities. Investigating the effect of data co-operatives on reducing the prominence of data extractivism through businesses, the author concludes that data co-operatives raise the awareness to issues of data gathering in urban environments in general but do not constitute a general trend reversal for equitable data gathering and governance.

3.2.6 Data Trust

Among the literature identified the topic Data Trust is also covered by the term Data Space and refers to an infrastructure where a trustee controls the distribution and use of the resource between multiple stakeholders like data subjects and data consumers.
Aziz Huq explains the institution of a public trust within which "an asset (…) is owed and managed by the state" (Huq 2021: 41) and gives examples of such public trusts controlling resources like public land controlled in the common interest. The author further proposes the establishment of a public data trust to protect the common interest of personal Data Protection from over-exploitation and privacy violation. Huq argues that municipal governments would constitute the best actors to create and operate such public trusts as their trustee because they are the closest link to the data subjects and have an interest in the equitable distribution of the resource themselves.
Paprica et al. (2020) draw experiences for the establishment of a public data trust regarding the case of health care data in Canada. The authors identify minimum requirements that such a trust should

respect mainly focussing on the governance of the trust (stated purpose, transparency, accountability, and adaptiveness) and the management (well-defined policies, Data Protection safeguards, and risk assessment and management), but also on training of the data consumers as well as the engagement of public stakeholders and data subjects (cf. Box 1 Paprica et al. 2020: 3).
In contrast, Houser and Bagby (2022) focus on the benefit the governance of data in a data trust could have for businesses and for the ease of global flows of data. They constitute specifically that diverging concepts of Data Protection in the United States and the European Union hinder flows of data and thus the economic viability of the resource for businesses. They theorise data trusts as *bundles of contracts* capable of addressing all legal concerns and making the data economy more accessible for smaller businesses and academics that cannot maintain an international data infrastructure themselves. Furthermore, they argue that such an international data trust could also create international traceability of the use of the resource for the data trustors (term often used for data subjects in the discourse of data trusts).
Lomotey, Kumi and Deters (2022) also address this concern of missing control over data distribution and usage areas for the data subjects and propose a Data Trust as a Service. They draft this as a private platform that allows the data subjects to understand and trace the usages of their data. As opposed to Huq (2021) proposing a governmental trustee for such a data trust, Lomotey, Kumi and Deters (2022) propose a private business as a trustee focussing on technical implementation and reliability questions. They outline possible use cases for example for water provisioning and traceability between the citizens as data subjects and the utility company as a data consumer.
Austin and Lie (2021) assess the concept of an Urban Data Trust that was part of the case study of Toronto's Sidewalk Lab, a well-known smart city project by Google parent Alphabet that was ended in 2020. Austin and Lie analyse the unclear aims of the concept that existed both as an open data portal in which data should be open by default, as well as an institution for collective data governance where the collective decides whom what data should be made accessible to. Besides this inconsistency in the objective of the urban data trust, they assign the failing of the concept to the incompatibility with Canada's Data Protection laws controlling personal data. Nevertheless, they assign potential to the concept of data trusts for urban data governance.
Further analysing the case, Scassa (2020) identifies three problems the Urban Data Trust concept faced; it was proposed as a reaction to criticism from the public over concern of data extractivism in the public realm, it re-defined personal data from the control of specific data subjects to be controlled collectively, and it was instituted by the main data consumer who would at the same time be the trustee of the data trust. Similar to the findings by Paprica et al. (2020) discussed earlier, Scassa (2020) concludes that the conceptualisation of an urban data trust has to start early, be discussed over defined data, and has to closely include the data subjects.

3.2.7 Open Data and Data Economy

References identified covering the topics Open Data and Data Economy discuss similar topics and are thus examined together in this section. The references mainly focus on economic value of data and the problems of inaccesibilities to data.

Lee (2021) inspects data-sharing among businesses but also from government data sources and argues that unnecessary restraints of sharing data result in economic inefficiencies for urban innovations. The author identifies three circumstances that hinder the data sharing economy: diverging standards and data formats, a missing legal and policy framework, and insufficient governance and management of data. These hindrances not only prevent economic exploitations of data but also the usage of data to serve the public good. Lee concludes all data should in the long term be considered a public good, but the path to that status will first lead to government data and later to big data being considered a public good.

Wernick, Olk and von Grafenstein (2020) by contrast argue that the problem is not necessarily data accessibility but rather invisibility and missing information about existence of available data. They argue that data intermediaries, i.e. middlepersons between data holders and data consumers, could alleviate this problem. They define two data governance models in which data intermediaries could play a strategic role. Data Clearinghouses, where an intermediary connects individual datasets to individual data consumers, and Data Pools, where the intermediary gathers datasets from multiple sources augmenting the value of the datasets to give it to data consumers (cf. fig. 1 Wernick, Olk and von Grafenstein 2020: 67).

While also advocating for better data accessibilities, Concilio and Molinari (2021) point to a subsequent problem of data accessibility, i.e. the limits to exploitability and problems of data quality in most open data. They study that missing data quality on the one hand stems from difficulties of maintaining and updating open data portals and thus argue for financial incentivising and regulating open data ecosystems. On the other hand, they argue that the lack of economic usage of data stems from questions of ownership and from private businesses withholding data and knowledge on how to assess this data from the public. To address these problems, they further investigate citizen science projects generating open data by their own means and conclude that further research in how to scale and support such initiatives is necessary.

Mercille (2021) also addresses the problem of missing data sharing from businesses and advocates for methods beyond voluntary sharing in smart city initiatives. The author argues that access to data means inclusivity of citizens as well as smaller businesses and data gathered by bigger businesses in smart city projects should therefore be made accessible for the general good. The author drafts a matrix consisting of four pathways of corporate data sharing; (1) State-driven, enforced data sharing with governments, (2) Corporate-driven, data sharing between businesses only, (3) Data Philanthropy, voluntary disclosure by companies, (4) Progressive, enforced disclosure of data. (cf. fig. 1 Mercille 2021: 6) Mercille subsequently calls for research in methods pursuing the fourth pathway.

Beckwith, Sherry and Prendergast (2019) similarly focus on the question of data access in smart city initiatives and the accompanied injustices between the stakeholders. They challenge the concept open data and explore the notion data as a commons to broaden the discussion beyond questions of mere accessibility and ownership of data generated in smart city initiatives. They argue as that the value of data is very objective and positive or negative uses of data depend heavily on the affected, the decision of who decides whose values are most important is best found by the community the data comes from. They thus conclude that a commons-based approach to data in smart city projects addresses questions of accessibilities, justice, and application of data best.

Lastly, the two references identified covering the topic Digital Platforms were also identified under the topics Data Governance (Micheli et al. 2020) and Data Co-Operativisms (Leclercq and Rijshouwer 2022) and were thus examined in these sections respectively.

3.3 Intermediate Discussion

The findings of the scoping review show the broad extent of the research topic and the diverse topics being discussed. However, despite the diverse backgrounds in fields of studies of the authors identified in the scoping review, precise concerns and topics have emerged that are shared among the specific disciplines and assessed from different standpoints.

The bibliometric analysis shows a clear advent of the topic data as a commons after the year 2018 (ref. fig.4). This can be brought into relation with the general rise of data gathering in urban environments but also with a sense for privacy and digital rights with the commencement of the GDPR in 2018. Among the disciplines contributing, it is noteworthy to observe that despite the overall dominance of legal scholars, scholars from the field urbanism and environmental studies appear to have propelled the academic discourse (ref. fig.5). Even though urban data concerns these studies immediately, this reflects an early sensibility for questions of digitisation and digital rights. Yet with a closer look at the individual references, urbanism scholars seem to propose solutions to the problems described, whereas other scholars focus on closer defining these problems. For example, Lupi (2019) proposing a City Data Plan to ensure accountability, or Leclercq and Rijshouwer (2022) endorsing co-creation platforms to address citizens' digital rights.

While the outcome of Data Governance as a prominent topic is unsurprising, given that the search terms specifically included "data governance", the prominence of the subsequent topics Digital Rights and Data Ownership is more interesting. This shows a strong criticism of the status quo of personal data gathering in urban environments and the importance of the topic in the overall digital rights discourse. Especially, as many of the subsequent topics like Data Co-Operativism or Data Trust equally cover questions of data stewardship, accessibility, and distribution.

Equally, the references covering the topic Data Governance address concerns over personal self-determination, data sovereignty, as well as access and inclusion themselves as well. The important dimensions in those matters the governance has to address are laid out (e.g. Zygmuntowski, Zoboli and Nemitz 2021) and equally the limitations governance faces (Prainsack 2019; Benfeldt, Persson and Madsen 2020). The attempt however to define what successful data governance would constitute quickly results in the conceptualisation of specific governance models (e.g. Lupi 2019) which contradicts the argument that the interaction of governance models allows the highest accessibility to procedural fairness (Smichowski 2019).

Furthermore, it is interesting to see how the same topics are discussed differently among the authors from different disciplines. For example, the dominance of privacy protection through laws is both criticised by legal scholars (Iaione 2019) as well as sociologists (Benjamin 2020) suggesting that questions of digital rights cannot be solved solely through top-down control but rather require a broad awareness and sensitisation in the public.

Similarly, the topic ownership is examined critically questioning the viability of the concept data ownership fundamentally (Hummel, Braun and Dabrock 2021) and pointing to ownership as diverse interactions between multiple stakeholders (Hicks 2022) that can be addressed from diverse perspectives (Drexl 2021). This fluent understanding suggests an importance of not only the actual ownership of data but rather the questions of accessibility, usages and the distribution of benefits generated from data, i.e. what data is collected and what it is used for. The concept of data as a commons reappears frequently as a way of addressing these economic injustices (Fia 2021; Park 2021).

Addressing the topic Open Data and Data Economy, the references focus on concerns of data quality and accessibilities. The insufficient interests of private businesses holding data to disclose their datasets (Mercille 2021) suggests an inefficient distribution of the resource as the enclosure prevents secondary usages of these datasets and the general usage for the common benefit. On the other hand, the insufficient quality of data openly available and also the knowledge of its existence (Concilio and Molinari 2021; Wernick, Olk and von Grafenstein 2020) points at both an overflow of data without sufficient meaning and the lack of digital literacy in the public.

Recurring topics in the literature are missing standardisation and data quality but also limits because of digital literacy and awareness in general suggesting that the topics have not yet reached full societal attentiveness and appropriate technical implementations. This refers to the second hypothesis that an empowerment of data subjects is necessary. Additionally to this, the dominance of the topics lack of data sharing and inaccessibility of private data silos present in the literature also support the first hypothesis of this study that the conditions of data production, rules of its use and the power relations between the stakeholders are not yet well defined.

In general, it is also interesting that references by urbanism and environmental scholars more often conclude with specific solutions (eg. Foth et al. 2021; or Choenni et al. 2022), whereas other disciplines examine the specific problems and concerns more closely. While most references stay broad and discuss topics generally, some of the references already point to specific methods of data governance. For example, by pointing to the role of data co-operatives in equitable data maintenance (Morell, Cigarini and Hidalgo 2021), data trusts as a controlled environment for sharing data (Houser and Bagby 2022), or with a checklist to ensure privacy rights of citizens in smart city projects (Wong, Henderson and Ball 2022).

However, many authors in the literature identified in the scoping review point to the lack of methods, call for a systematisation of methods, or a bigger focus on equitable governance of data in smart city projects (eg. Micheli et al. 2020; or Delcroix 2017). The question of specific methods for an equitable governance of data existing in real world usage thus becomes pressing. Pursuing this question, the interview findings will reveal important insights to further answer the research question of this study in the next sections.

4. Interview Findings

The findings of the scoping review have shown that diverse concerns and topics are at play and that there is not yet an overview of methods surrounding an equitable governance. Therefore, the aim of the expert interviews is to identify methods used in practical examples and also to understand the limitations and concerns cities are facing. The interviews have been transcribed and assessed qualitatively and the results of this analysis will be shown in this chapter.

4.1 Coding Result

During the transcription, it already becomes clear that besides the different kinds of methods, also the different aims they pursue seem to be relevant. Therefore, besides the deductively set codes general discussion and the types of methods, the coding process induces additional code families and thus codes emerge. The matching and organisation of the codes groups relevant codes together and generalises the statements to making them more comparable among the different interviews. The ensuing code system can be split into three groups with subsequent categories, the numbers stated refer to the positions coded in the interviews that can refer to multiple codes:

- General Discussion of Data as a Commons (n = 138)[3]
    - Limitations/Difficulties (n = 72)
    - Tragedy of the Data Commons (n = 29)
    - Understanding of Data as a Commons (n = 19)
    - Ownership of Data (n = 18)
- Methods of Governance (n = 194)[3]
    - Legal Methods (n = 30)
    - Input-Side Methods (n = 31)
    - Output-Side Methods (n = 116)
    - Design Principles (n = 21)
- Aims Methods (n = 131)[3]
    - Data Protection (n = 43)
    - Creating Awareness (n = 63)
    - Collecting Consent (n = 18)
    - Enabling Interaction (n = 57)

All the methods mentioned can be categorised under the four categories Legal Methods, Input-Side, Output-Side Methods, and Design Principles, while for the aims four concepts emerge. Data Protection, as in protecting digital rights and the handling of personal data, Creating Awareness, and Collecting Consent for data gathering, as well as Enabling Interaction with the data gathered for the data subjects involved.

[3] Some positions in the interviews refer to multiple codes; one position may refer to both Data Protection and Creating Awareness. Thus the number is not the sum of all positions of the subcategories.

Further, the different aims are split into whether the interviewee considered it to be already pursued by specific methods today, or whether there exists a gap of methods, i.e. there is no satisfying method pursuing this specific aim. For example:

- Collecting Consent (n = 18)
  - Gap (n = 6)
  - Pursuit (n =12)

Some codes emerging during the coding processes that were only mentioned a few times or could be transferred to other code topics are subsequently not considered. Based on the co-occurences of the codes and amount of mentions, the topics will be evaluated and explored in the following sections in general and for the specific cities thereafter.

4.2 General Findings

In this chapter the results of the qualitative analysis of the interviews are shown highlighting the relevant topics emerging in the interviews. The results are first presented regarding the findings from all cities focussing on differences between city representatives and public representatives in general. Thereafter, the results for each individual city is shown and the main methods relevant in these cities are identified.

4.2.1 Understanding of Data as a Commons

Different ways of understanding the term data as a commons can be identified: the first and present understanding in this study is the understanding of data as a CPR, the second understanding describes the value that data creates for the society (Data for the Common Good), the third understanding is that only Open Data can be considered a common good. Whereas one particular understanding of data as a commons is an understanding as a Commons Ecosystem in which only data within that defined ecosystem is considered a common good.

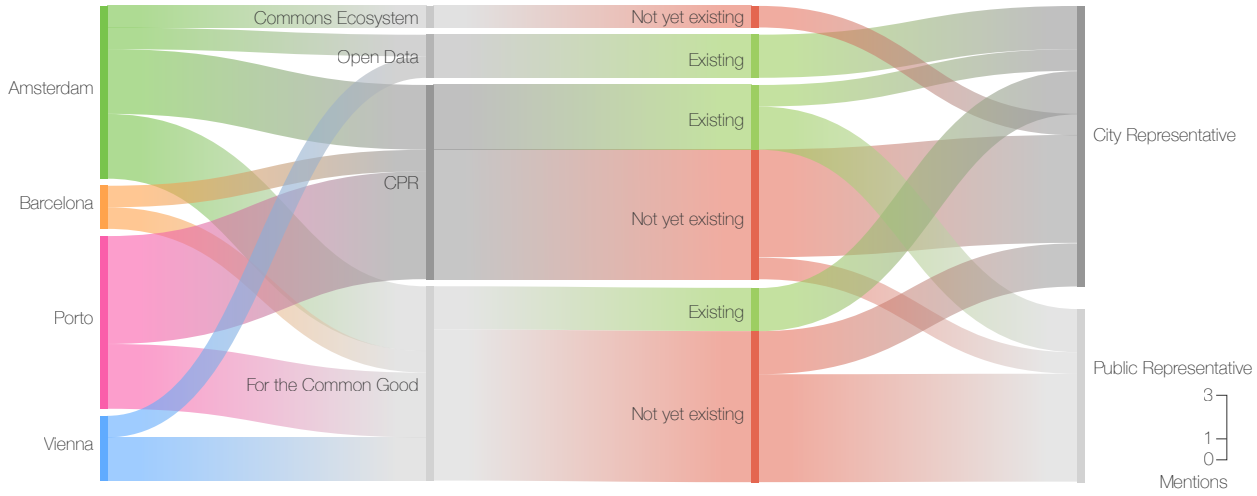


Figure 6: Different Understandings of Data as a Commons*

The two ways of understanding data as a commons as in a CPR and Data For the Common Good seem to be equally represented in all interviews (both n=9), whereas Data Commons as Open Data was mentioned twice and the concept of a Commons Ecosystem only once. The second understanding, Data for the Common Good, is the only understanding that seems apparent in all cities. However, as the number of occurrences from the city perspective is directly dependent on the number of interviews conducted in the individual cities, the representation does not reflect that the interviewees in Amsterdam and Porto are more opinionated on this question than stakeholders in Barcelona and Vienna. Rather, only the relative distribution for each individual city is decisive. Here, the perception of Data for the Common Good plays the biggest role for interview partners in Vienna, whereas the understandings data as a CPR and for the Common Good was equally represented in interviews with stakeholders in Amsterdam. Surprisingly, the understanding data as a CPR was absolutely and relatively represented the strongest in the interviews with stakeholders in Porto, although the City of Porto did not include the concept data as a commons in their digital agenda as opposed to Amsterdam and Barcelona. However, the different understandings are not mutually exclusive, e.g. an understanding of data commons as Open Data does not rule out an understanding of data for the Common Good.

Further, the positions towards the different understandings of data as a commons become interesting. While the understanding of Open Data as a commons is completely positive and a data Commons Ecosystem is perceived as an aim not yet reached, data as a CPR and data for the Common Good are perceived more controversial by the interviewees. Predominantly, the status of both understandings is considered to be not yet fulfilled, i.e. data is perceived as not maintained as a CPR today and data is perceived as not used for the common good today. This reflects the perception that a lot of data is unopen and in control of private businesses that own the data.

On the other hand, the difference in general perception between city representatives and the outside view on the city shows a general disparity; While for the city representatives the distribution is approximately equal with eight critical opinions and seven positives, the distribution for public representatives has five critical attitudes and only two positive expressions. Furthermore, this disparity is even more distinct if looked at the individual understandings; while the perception of the status of data as a CPR is mainly critical for city representatives, public representatives express primarily positive sentiments. In contrast, usage of Data for the Common Good was perceived as both existing and pursued. Whereas public representatives expressed a gap in the use of Data for the Common Good.

With these apparent discrepancies between understanding of data as a commons and also the status of the different understandings, it is important to learn about the different limitations and difficulties in the area of data governance that are perceived in the cities.

4.2.2 Limitations and Difficulties

Depending on the individual backgrounds and areas, the interview partners responded with diverse problems they are facing in data governance and data gathering inside the city government and in the public with data gathering projects.

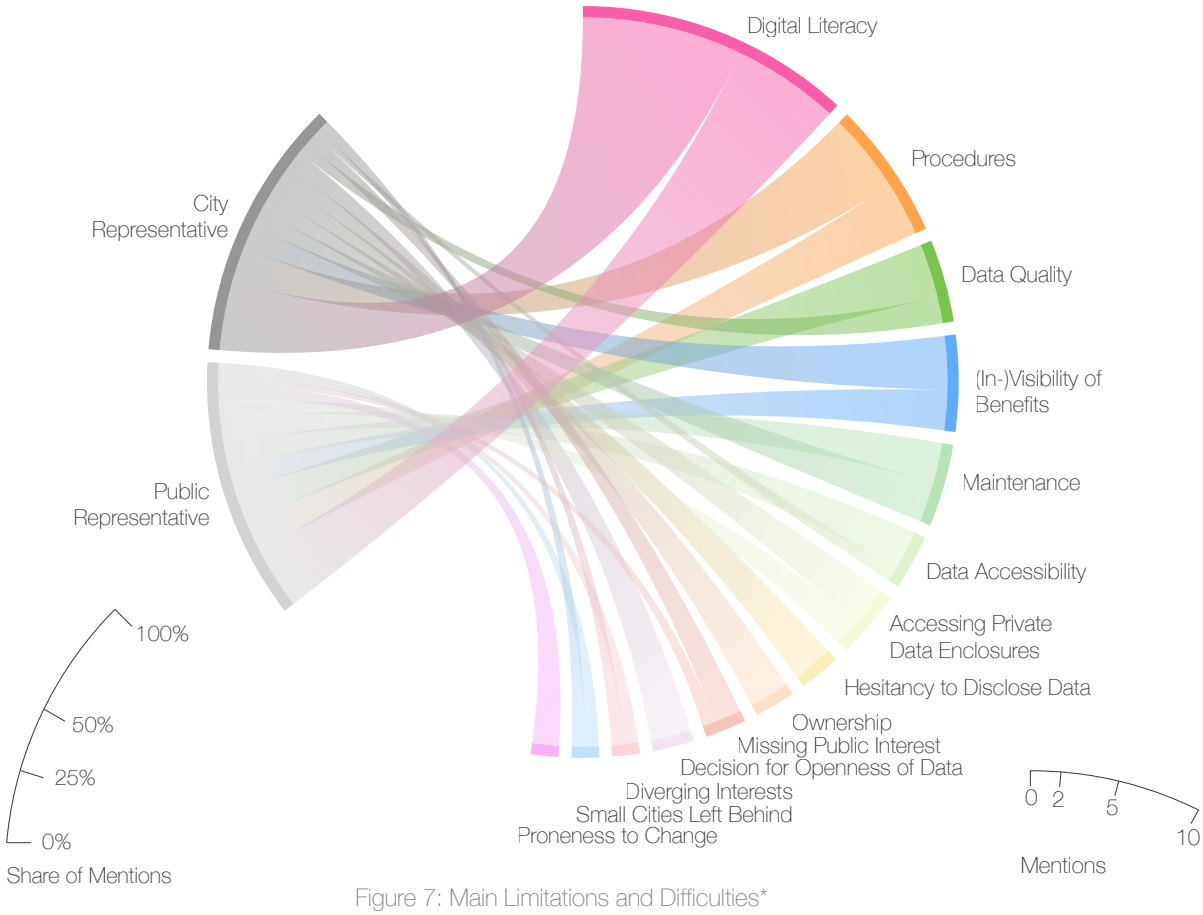


Figure 7: Main Limitations and Difficulties*

In general, there have been more mentions of difficulties and limitations from the side of city representative, but more specific voices from the public representatives. Among these limitations and difficulties, digital literacy and procedures appear to be the most persistent and are equally voiced by city representatives and outside the city governments. This means the competencies not only inside the city government but also of the general public to know how to interact with data and questions of digitisation and the problematic of gridlocked regulations or lack of clear rules respectively. Specifically for the difficulties of data quality, maintenance, data accessibility and the issue of proneness to change, the distribution seems to shift to a more critical assessment from the outside view on the city government. Here, the concern for missing quality for specific use cases and the methods of data gathering and the issues of availability of fitting data is voiced. On the other hand, questions of data ownership, missing criteria for what data should be open data, and a hesitancy to disclose data seem to be a bigger concern for public representatives, as only they voiced these concerns. The difficulty of making the benefits visible to the general public has been addressed equally as often by both public and city representatives. Solely the issue of proneness to change within the city governments was voiced by the public representatives only.

4.2.3 Tragedy of the Data Commons

Following Garret Hardin's (1968) influential work, *The Tragedy Of The Commons*, one of the interview questions was searching for a potential tragedy of the *data commons*. Hardin (1968) argued that an uncontrolled CPR would inevitably lead to an overuse and thus a depletion of the resource, because every consumer does not have an interest in lowering their own use of the resource unilaterally. Being a non-rival good, the resource data cannot be diminished through its use as one use of the resource does not compromise another. However, Greco and Floridi (2004) have argued that the tragedy of *digital* commons could constitute in a digital divide, where the reliance on data for decision making would leave certain people out of the discussion as they do not have the education and lack the tools to assess data and data gathering issues. Bloom (2013) argues that this tragedy is constantly imminent as private data collectors making data inaccessible and uninteroperable but renders this danger of a tragedy of the data commons far from inevitable if managed in a correct way.

Faced with the question, what a tragedy of the data commons could look like, the interview partners in the individual cities expressed diverse opinions.

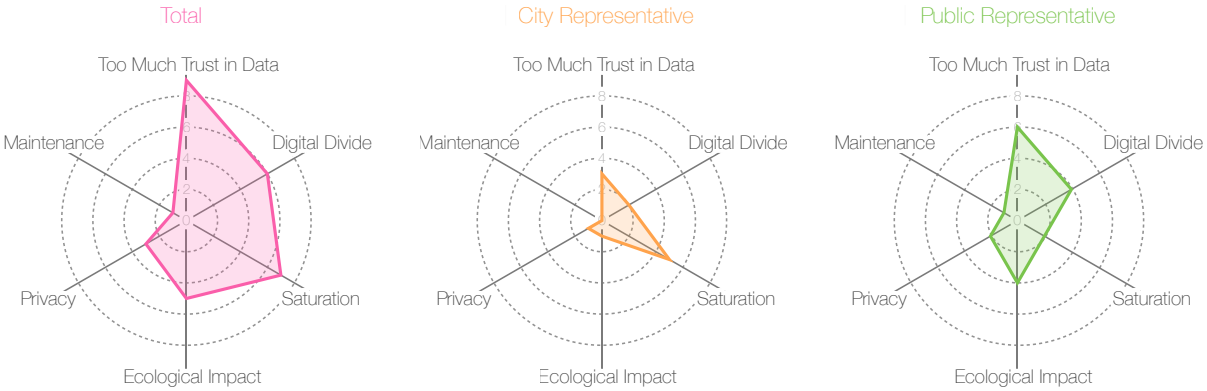


Figure 8: Tragedies of the Data Commons*

In general, public representatives voiced more opinions of tragedies than city representatives. Despite the prominence of the digital divide as a tragedy of the data commons in the literature, the main concept in the answers is the idea of having too much trust in data. This means relying too much on data-driven decisions without questioning the sources and meaningfulness of the data. Another tragedy mentioned could be in the saturation of data. This has been stated as an overflow of data, meaning that a vast amount of data is preventing a meaningful usage of it. This can be particularly striking if datasets suggest contradicting facts because of different ways of data gathering or different focusses. Further, this saturation could also render a tragedy through the meaninglessness of the dataset, the tragedy that data is gathered, distributed, and made accessible without any use in the end. Another dimension to this concept of saturation is the idea of economic depletion through substitutability, i.e. if a dataset can be easily replaced or gathered in another way, the economic value of this individual dataset vanishes. An example of this could be a company that today earns its money by accumulating and selling mobility behaviour data that could also be gathered by all individuals sharing their data and making it accessible.

In the interviews, the tragedy through a digital divide relates to the limitation of digital literacy, i.e. the risk that if we become more dependent on data for urban development decision-making, people who do not know how to understand data or do not have the tools to use the data are left behind and excluded from the discussion. As data is considered a non-rival good, it does not have to be duplicated to be used over and over again and thus does not require more resources. However, the infrastructure itself that stores and makes data accessible requires growing amounts of resources. Here, one tragedy mentioned lies in the ecological impact that data, or more specifically data storing and distributing. Through data centres and infrastructure in the physical world data translates to energy consumption and pollution and could thus constitute a tragedy of the data commons. One further tragedy mentioned in the interviews is the tragedy through loss of privacy because of data gathering and the subsequent psychological impact this could have on data subjects. This means that through ubiquitous data gathering and evaluation, personal self-determination could get lost if the data is used to influence the data subject's behaviour in a certain way or if the data is made publicly available. Even if only stated once, the problem of maintaining the ecosystem to provide data could be a tragedy of data commons. If datasets today are published, the effort of keeping the data up to date and allowing enough accessibility to it could become a tragedy of a commons ecosystem tomorrow as this requires human and infrastructural resources.

Among the possible tragedies stated, public representatives primarily stated concerns of too much trust in data and the digital divide, whereas city representatives primarily focussed on the question of data saturation. The different concepts presented show that even if data can be used numerous times without depleting, a tragedy could still occur. However, as these concepts only reflect the spontaneous answers the interviewees gave, the frequency does not reflect an importance of one concept over another.

4.2.4 Aims and Methods

With digitisation being a new reality city governments have to take on, the question of their role in the process of digital rights being also a topic frequently discussed. Beyond concepts of bottom-up grassroot development of data gathering or top-down management through data, the role of the city government is predominantly described as an institution that empowers its citizens to a self-determined development of the city through digitisation projects. Be it by incentivising participation through organising events, providing financial aids to data cooperatives, or by regulating data control within private business. Different methods were suggested and discussed city governments can take to enable equitable data flows in the city.

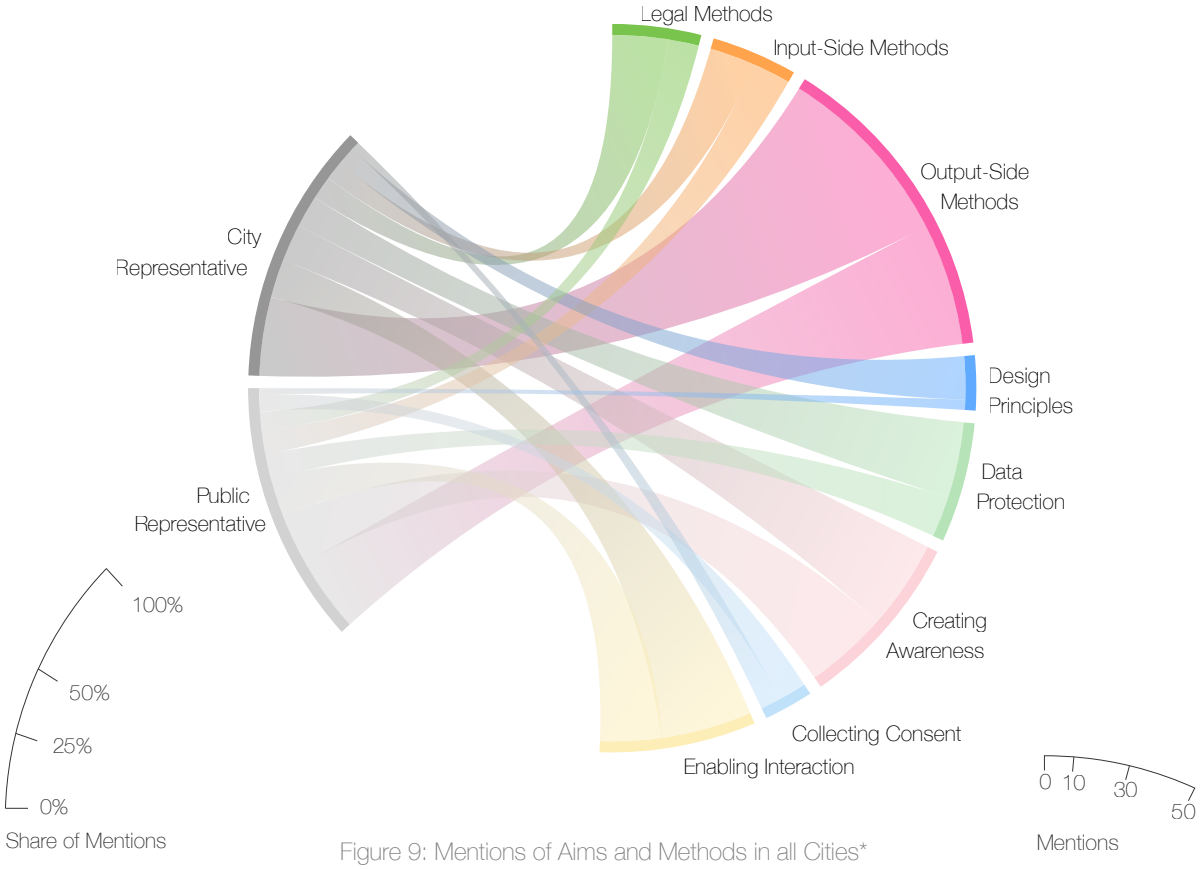


Figure 9: Mentions of Aims and Methods in all Cities*

Figure 9 shows the distribution of mentions of types of methods and aims among city representatives and public representatives from all four cities. In general, the mentions of both types of methods and aims are spread evenly between city representatives and public representatives without any mentions only stated by one particular group. Among these different types of methods, predominantly Output-Side Methods were mentioned both by city representatives and public representatives. Similarly, the aim Creating Awareness has been mentioned the most in total. However, from the city representatives' perspective, the aim Enabling Interaction has been mentioned more often than Creating Awareness. Likewise, Design Principles have been mentioned more often by city representatives than by public representatives. Even if the aim Collecting Consent has been mentioned the least in total, it has been mentioned more often by public representatives than by city representatives.

Furthermore, there were not only mentions of an aim being already pursued by a specific method, but also of aims not having been properly addressed yet. Here, another image emerges.

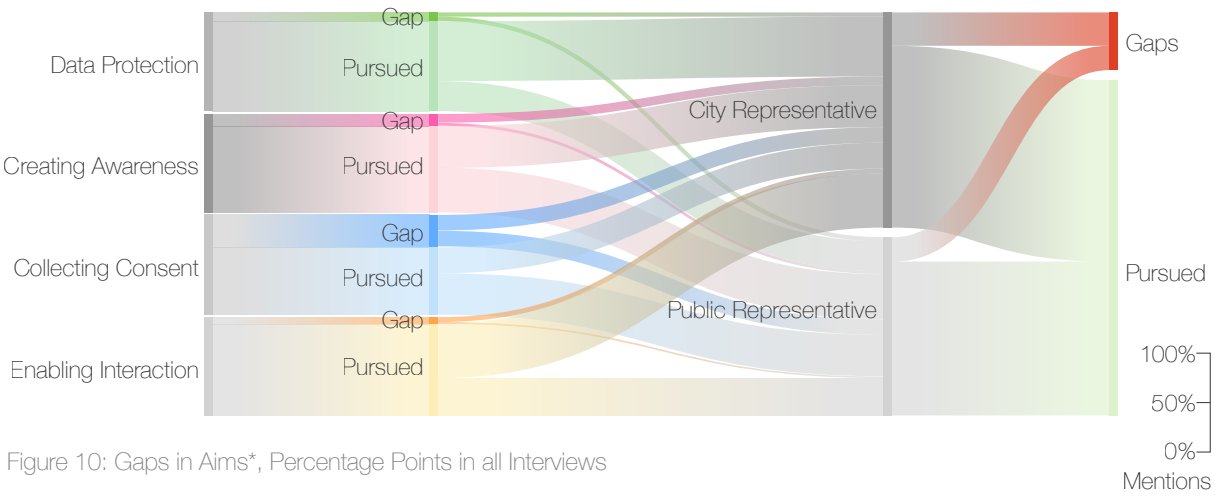


Figure 10: Gaps in Aims*, Percentage Points in all Interviews

Among the mentions of aims, the majority of mentions were given in the context of a specific method that pursues an aim. However, every eleventh mention of the aims Data Protection and Enabling Interaction, every eighth mention of Creating Awareness, and every third mention of the aim Collecting Consent have been pointing to a gap in the pursuit of this aim. Considering that the aim Collecting Consent has been mentioned the least (ref. fig. 9), this is especially interesting. The mentions of gaps in this aim came equally from city and public representatives, whereas the positive mentions of this aim came mainly from public representatives. Conversely, the aim Creating Awareness was as often mentioned in the context with a specific method by city representatives as by public representatives, but gaps were mentioned more often by city representatives. For both aims Data Protection and Enabling Interaction, the positive mentions of these aims came more often from city representatives.

Seen from the perspective of mentions of gaps and pursued aims in general (fig. 10 from the right), the mentions of distribution seems generally equal, with more mentions from city representatives in general.

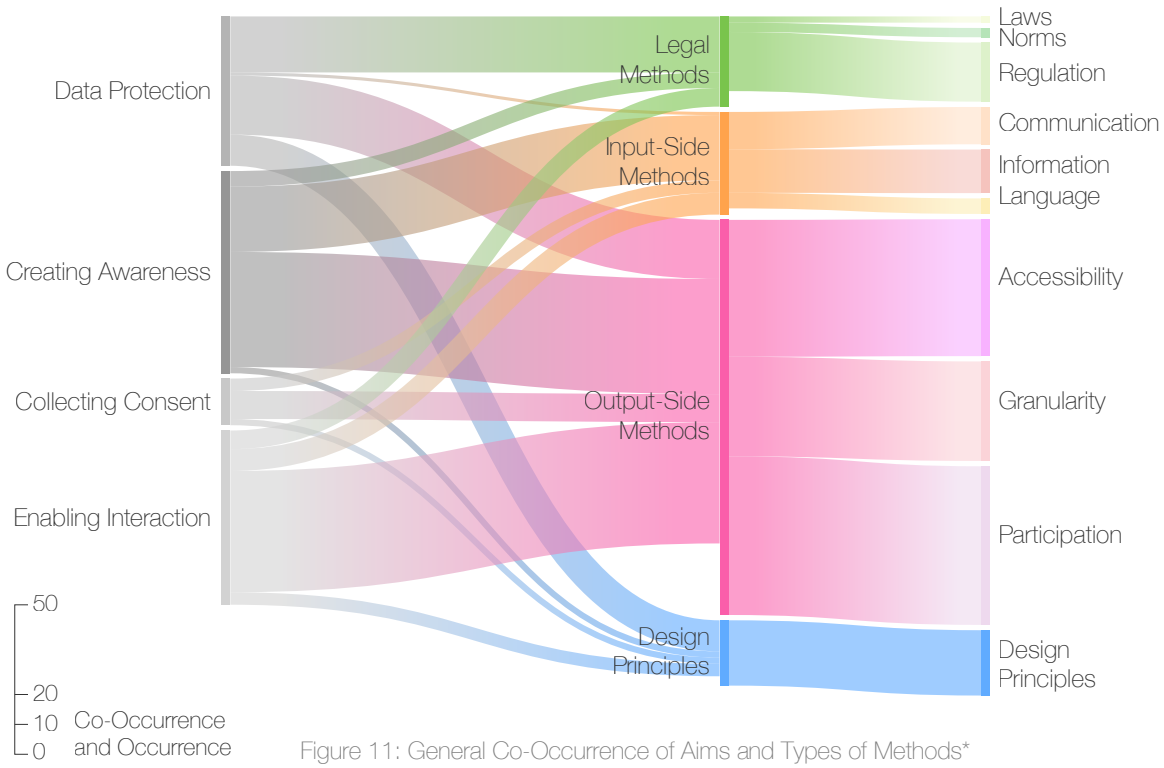


Figure 11: General Co-Occurrence of Aims and Types of Methods*

Figure 11 shows the co-occurrences of aims and types of methods, i.e. what aims were formulated in the context of which types of methods. It also shows which categories of methods were mentioned the most in all cities combined. Here, Output-Side Methods and specifically methods of Data Accessibility and Participation play the biggest roles. Granularity refers to methods that change the degree of informational value in datasets, for example by aggregating data. In general, it seems like Legal Methods are primarily addressing the aim data Data Protection, Input-Side Methods the aim Creating Awareness, and Design Principles the aim Data Protection as well. Output-Side Methods have been primarily mentioned in the contexts of Creating Awareness and Enabling Interaction. However, from the perspective of the aims, Output-Side Methods have been mentioned in half or more of the cases in the contexts of the aims Creating Awareness, Collecting Consent, and Enabling Interaction, and a third of the time in the context of Data Protection, thus addressing all the aims that were stated. Specifically for the aim Collecting Consent, Output-Side Methods play a bigger role than Input-Side Methods according to the interviewees.

The general breakdown of the aims and methods show the general dominance of Output-Side Methods both in the amount mentioned in general and also in the context of the aims mentioned during the interviews. Nonetheless, Legal Methods, Input-Side Methods, and Design Principles have all been evenly mentioned as well. On the other hand, the aims Data Protection, Creating Awareness, and Enabling Interaction have been mentioned similarly often, while Collecting Consent appears to play a minor role.

On the background of the general analysis, the interviews will be evaluated for the cities individually in the next chapter looking both at what city representatives and public representatives mentioned as well as the specific dependencies of aims and methods for the cities.

4.3 Individual Case Findings

In this section the findings for the individual cities are shown. For each city, the general distribution between city representatives and public representatives of what aims and methods they have referred to is visualised and examined. Then, the share of what types of methods and the methods specific for the individual city are closer analysed. The most important methods identified in the interviews are reviewed more closely in section 6.2 Review of Specific Methods.

4.3.1 Amsterdam

The interview partners in Amsterdam comprised of four city representatives - in the digital ethics field, in the specific data as a commons debate, in the technical implementation of data gathering and an open data portal representative - as well as two interview partners from public interest groups - the Waag Society[4], a lab for societal discussion of technology and the Digital Rights House[5]. The interviews were conducted mainly between June 13th to June 16th and one interview on June 28th 2022.

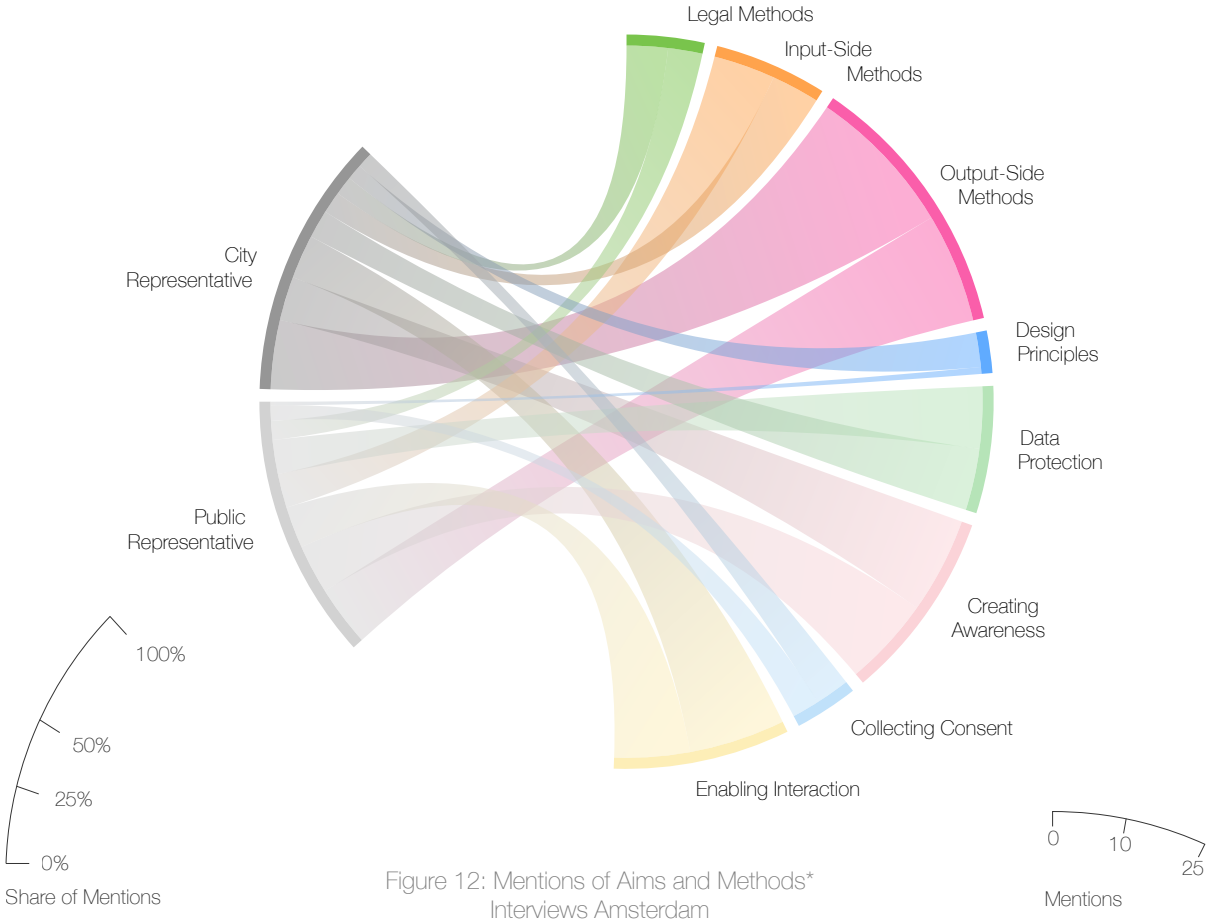


Figure 12: Mentions of Aims and Methods*
Interviews Amsterdam

The different types of methods and aims they pursue being mentioned in the interviews show that in Amsterdam the aims Data Protection, Creating Awareness and Enabling Interaction seem to be prominent. Output-Side Methods appear to be equally dominant. However, compared to the general

[4] waag.org last accessed: 22.08.2022
[5] digitalrightshouse.org last accessed: 22.08.2022

amount of mentions of methods and aims (fig. 9), Legal Methods and Input-Side Methods are mentioned relatively more often than in all four cities combined. Especially regarding Input-Side Methods, these have been more often mentioned from the side of public representatives. Similarly, the aim Data Protection has been more often addressed from the public representatives than from within the city government.
Apart from these two particularities, the amount of mentions of methods and aims is very homogenous among city representatives and public representatives.

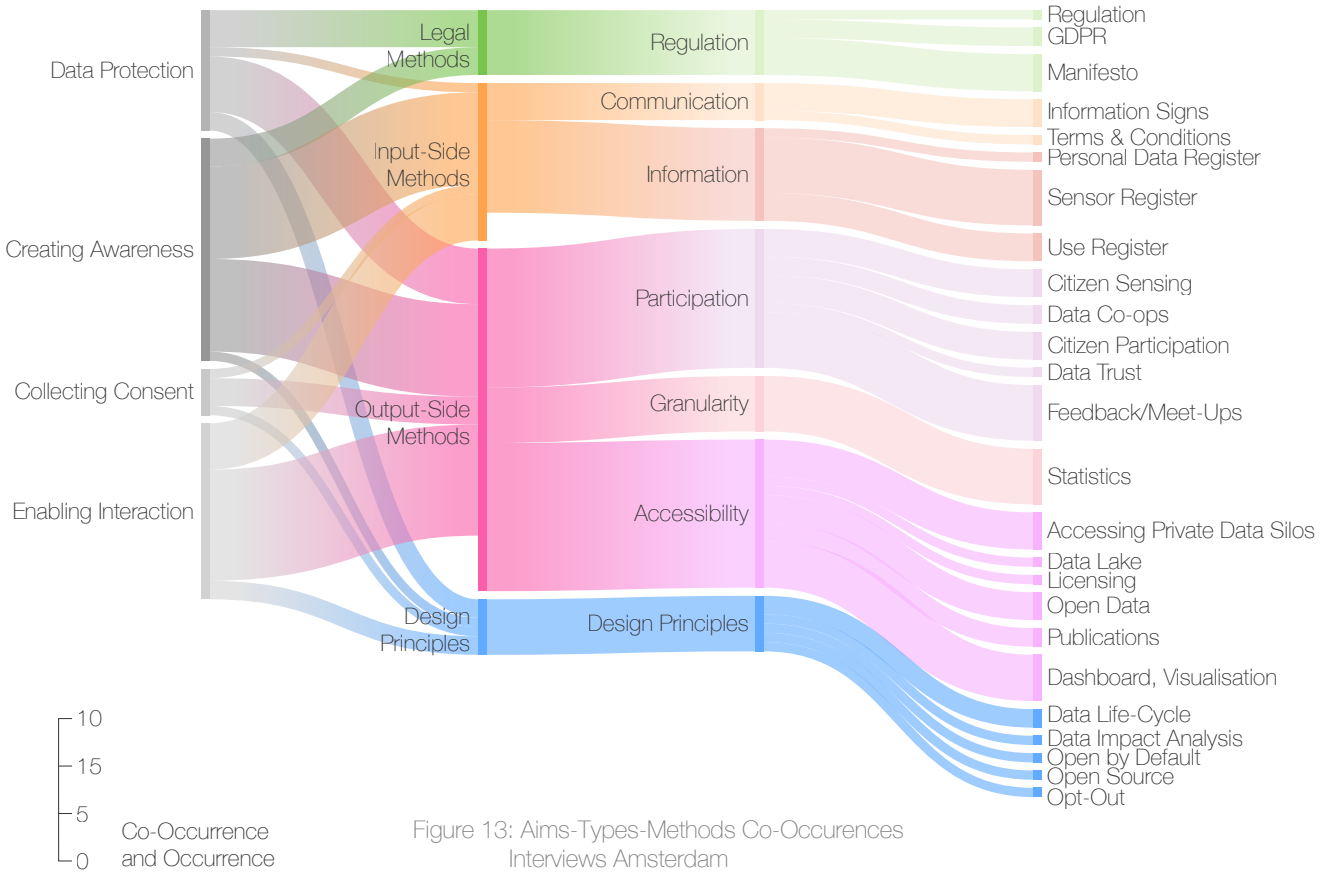


Figure 13: Aims-Types-Methods Co-Occurences Interviews Amsterdam

For the case of Amsterdam, Legal Methods are only discussed in the context of the aims creating awareness and foremost Data Protection, whereas Input-Side Methods are mainly considered for Creating Awareness about the gathering and enable interaction with the data gathered. Output-Side Methods seem to be discussed for all four aims. From the perspective of the aims, the majority of mentions of aims occur while Output-Side Methods are being discussed. Here, methods of Participation and Accessibility to data play a major role. Through the principle Open by Default and Data Impact Analysis, the method type Design Principles occur largely in the context of the aim Enabling Interaction and Data Protection.
Among the methods discussed, specific methods for Amsterdam become apparent. For example, Manifestoes were discussed that can be understood as a guideline or target formulation in data governance for city governments by which their work can be checked and monitored. Here, the *TaDa manifesto*[6] was discussed that formulates six core principles for responsible data gathering in cities. Further, Information Signs in the public sphere play a role that communicate not only the

[6] tada.city last accessed: 22.08.2022

gathering but also further information on the use of data. The project *Shuttercam*[7] by the Responsible Sensing Lab tested in Amsterdam allows users to shut off cameras for a defined time. Another often mentioned method is the Sensor-Register[8] the city of Amsterdam has implemented. Here, all publicly owned sensors and algorithms used are documented on a map to create transparency and awareness about data gathering. In the environment of the open data portal, the city government also holds a weekly meetup, *DataLab*[9], as a podium for uses of open data in Amsterdam.

4.3.2 Barcelona

In Barcelona, there was one city representative working on the Open Data Portal of Barcelona, and two public representatives; one person from the Digital Commons Research Network[10], a research institute at Open Universtiy of Catalonia, and one from the Institute for Advanced Architecture of Catalonia[11] (IAAC), an education and research centre focussing on digitisation in the built environment. The interviews were conducted between June 21st and 23rd 2022.

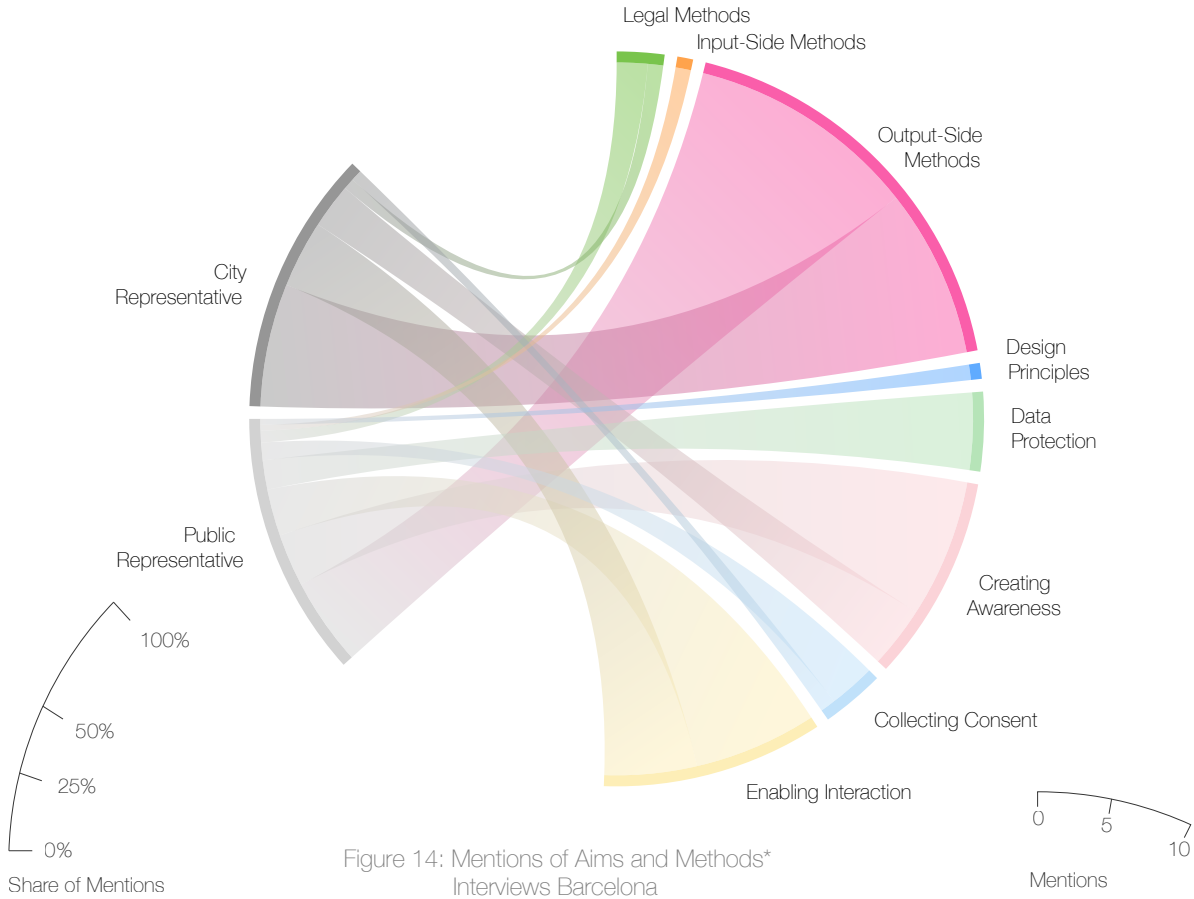


Figure 14: Mentions of Aims and Methods*
Interviews Barcelona

On the background of Barcelona's citizen-empowerment and -involvement approach, the dominance of Output-Side Methods becomes clear. Interestingly however, even though the

[7] responsiblesensinglab.org/projects/shuttercam last accessed: 22.08.2022
[8] sensorenregister.amsterdam.nl last accessed: 22.08.2022
[9] amsterdam.nl/datalab last accessed: 22.08.2022
[10] dimmons.net last accessed: 22.08.2022
[11] iaac.net last accessed: 22.08.2022

mentioning of Output-Side Methods is comparatively higher than for the general distribution (ref. fig. 9), the distribution of mentioned aims is less tilting with the aims Data Protection and Collecting Consent still substantially relevant.
While the city representative mentioned primarily Output-Side Methods and focussed on the aims Enabling Interaction and Creating Awareness, the public representatives also engaged in the discussion of Collecting Consent and Data Protection. However, it seems like Input-Side Methods and Design Principles have been playing a subordinate role in the interviews in Barcelona.

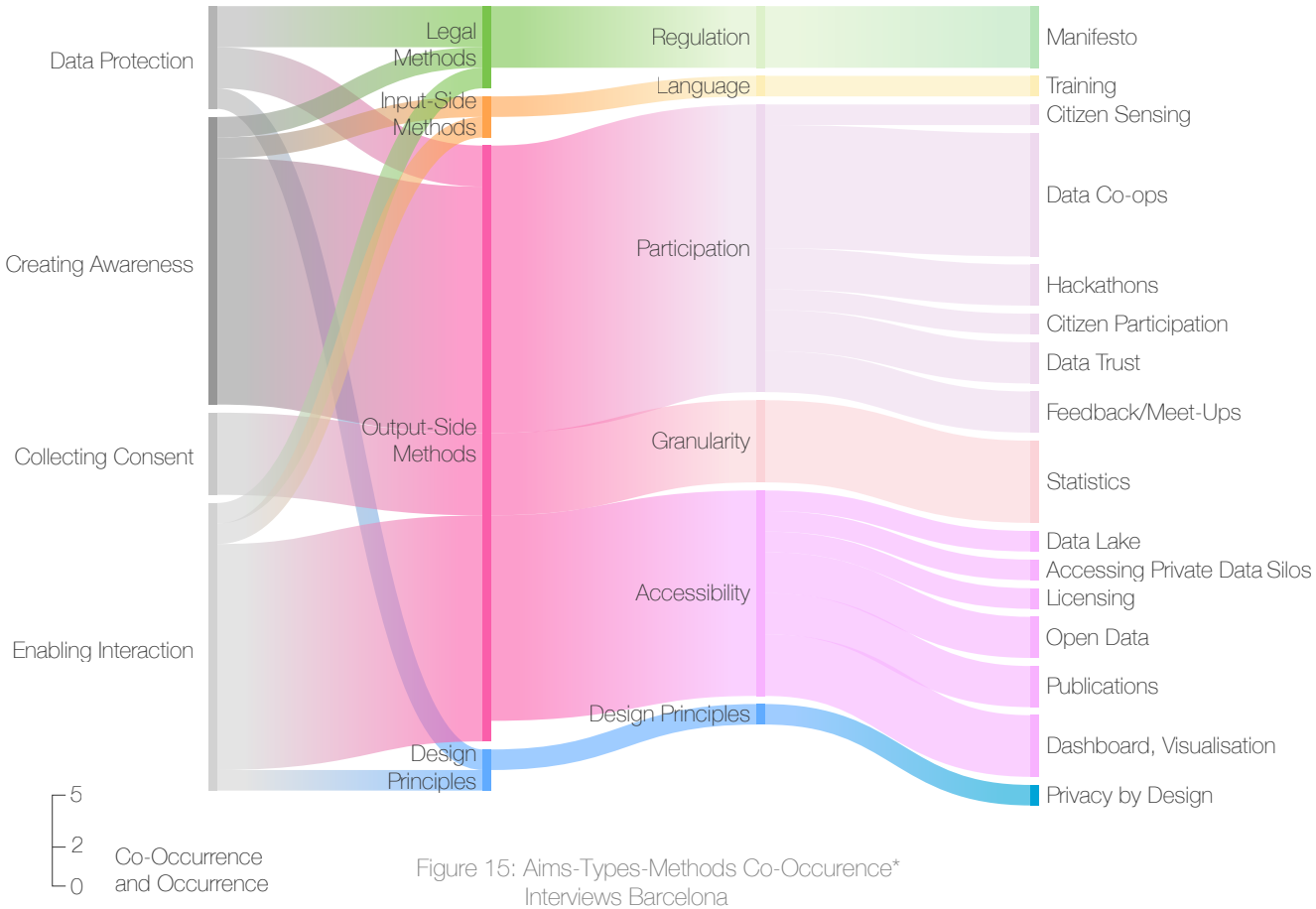


Figure 15: Aims-Types-Methods Co-Occurence*
Interviews Barcelona

The dominance of Output-Side Methods is also visible in the interrelation with the aims mentioned in the interviews. Here, these types of methods are also addressing all the aims in the interview discussions. However as opposed to Amsterdam, there is a stronger concentration on the aims Creating Awareness and Enabling Interaction. Interestingly, also the aim of Collecting Consent is discussed exclusively in the context of output-side participation.
Specific methods identified and discussed in the interviews in Barcelona are for example Data Co-Operatives that have been mentioned six times as a concept of shared data governance among citizens and an example *Salus Coop*[12] that manages health data in a collective way. Citizen sensing and citizen participation however have been playing a significant role, with projects like the FabLab Barcelona[13] initiated by the Institute for Advanced Architecture of Catalonia. Another particularity is the making available of acquired datasets from companies, accessing these private data silos either

[12] saluscoop.org last accessed 22.08.2022
[13] fablabbcn.org last accessed 22.08.2022

by buying the datasets or by implementing regulations that require the sharing of specific datasets. The datasets are either fully made available or are simplified to a lower degree of granularity using statistical methods to meet licensing requirements. Similarly, the data available in the open data portal is not only accessible as data in different formats or through an API, also the information contained in the data is communicated through visualisation, dashboards, or publications to reach more citizens than only the ones who know how to work with data. Another method shortly addressed is the method Data Lake, a technical infrastructure for storing all data and data types centrally with different access rights.

4.3.3 Porto

There were four interviews in Porto with three city representatives, one of them working in the technical implementation area of digitisation projects, one in the civic engagement area, and one interview with two interviewees regarding Data Protection, as well as one interview with a public representative who is a member at the architecture faculty of Universidad da Maia and member of the Open and Agile Smart Cities[14] network. The interviews were conducted on June 27th, 28th and July 17th 2022.

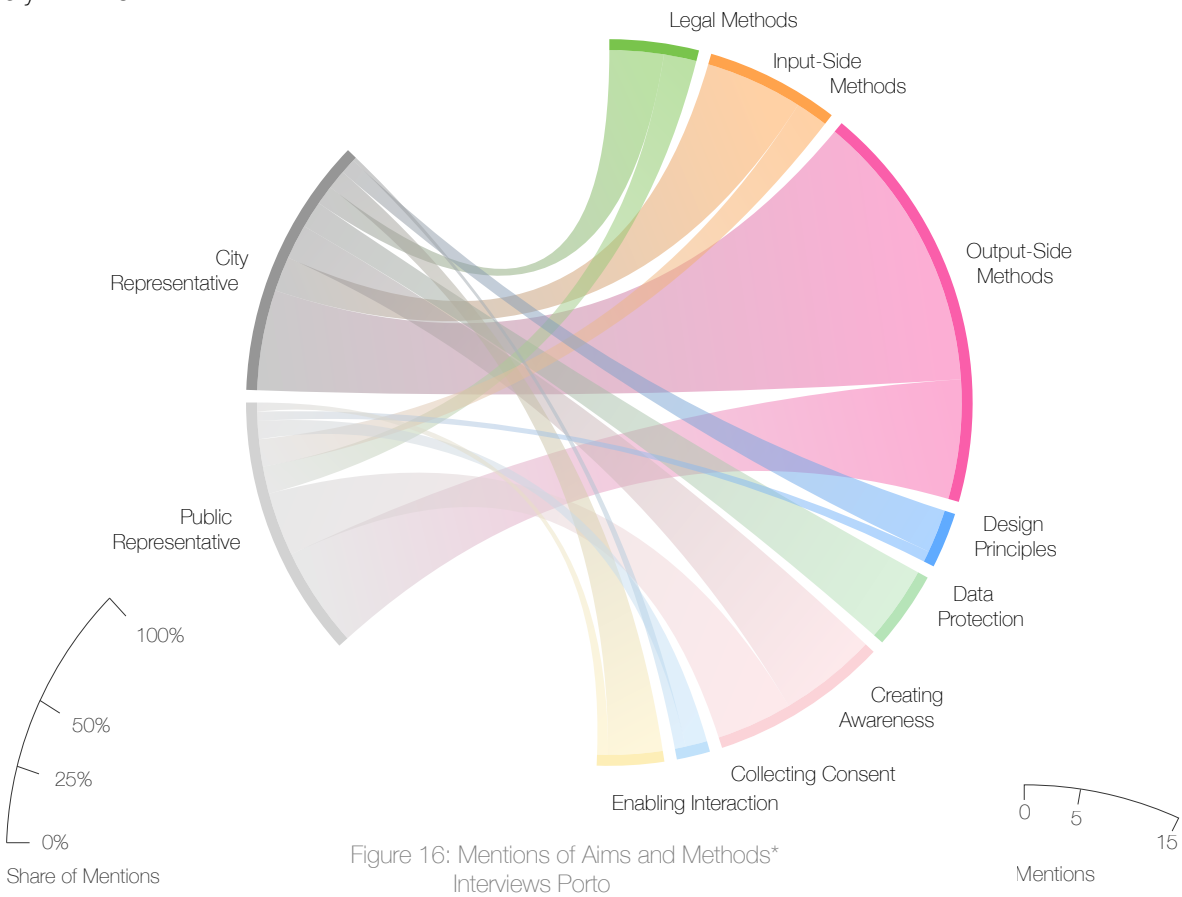


Figure 16: Mentions of Aims and Methods*
Interviews Porto

Throughout the interviews in Porto, more methods were mentioned than the aims they pursue. Conversely, among the methods the different types were discussed in an equal distribution among the city representatives and the public representative; there appears to be no tendency towards one

[14] oascities.org last accessed 17.08.2022

type of method from the city representation or outside view. Here, Output-Side Methods also play a major role, while Input-Side Methods were likewise mentioned frequently.
As opposed to the interviews in Barcelona, the aim Creating Awareness seems to play a bigger role in the interviews in Porto. This aim is also the main aim stated by the public representative, whereas the aims Enabling Interaction and Collecting Consent have been less mentioned by them. By contrast, the city representatives in Porto have been stating the aims Creating Awareness and Data Protection nearly the same number of times, while also stating the aim Enabling Interaction more often than the public representative.

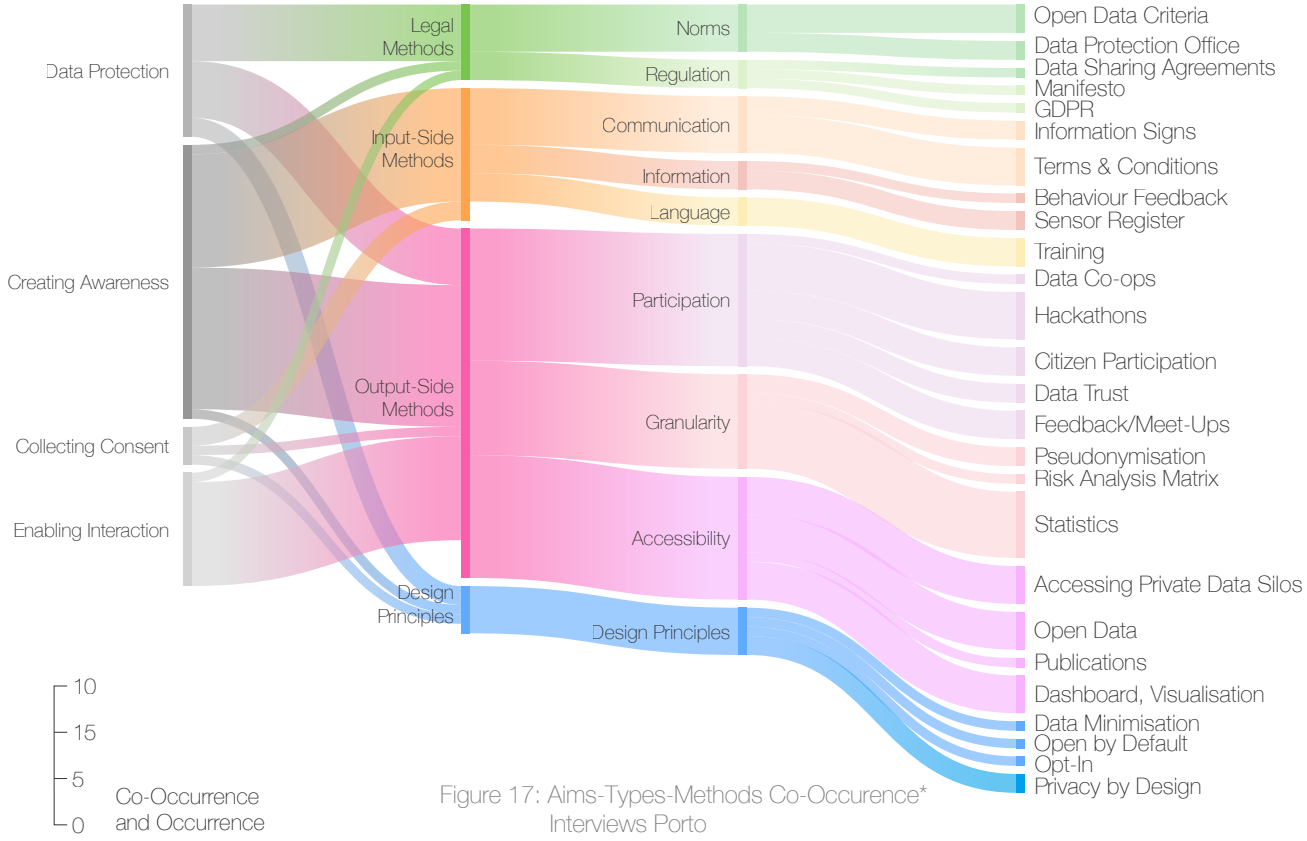


Figure 17: Aims-Types-Methods Co-Occurence*
Interviews Porto

While Output-Side Methods have been mentioned in the context of the aims of Data Protection, Creating Awareness and Enabling Interaction, Input-Side Methods have been primarily discussed in the context of Creating Awareness. Here, primarily the information about data gathering through Terms and Conditions in applications and Training Programs for city employees have been discussed. Even though Terms and Conditions are a classical method of Collecting Consent, the method has also been discussed in the context of Creating Awareness.
Some of the methods discussed have been mentioned in the context of a specific project, the wifi network on public transit in Porto provided by the company Veniam[15]. Here, users are provided with internet and provide their data in return. This happens by saving the specific mac addresses of the devices connected to the Wi-Fi, i.e. the unique identifiers of devices constituting personal data. Meeting the Data Protection regulations of GDPR, consent is collected through the terms and conditions that every user has to accept before using the service. For further processing of the dataset, the mac addresses are pseudonymised, i.e. given a new name so the mac address is not

[15] veniam.com last accessed 17.08.2022

directly identifiable anymore. The data is then used among other uses to visualise the degree of crowdedness on buses, i.e. publicly available in a lower degree of granularity, as well as to optimise routes, i.e. only available within the public transport company. Another method the city of Porto has experiences with are Hackathons[16]. Usually annually, the city organises an event where participants are given a design challenge to solve using open data and specific datasets made available for the event. In the context of health data, a Risk Analysis Matrix was discussed that analyses the risk of identification in a specific dataset using an algorithm and blocking access if the risk is deemed too high.

4.3.4 Vienna

The three interviews in Vienna were conducted on June 9th, 10th and July 19th 2022. Participants were two city representatives, the open data coordinator and the digitisation coordinator, as well as a public representative focussing on AI implementations.

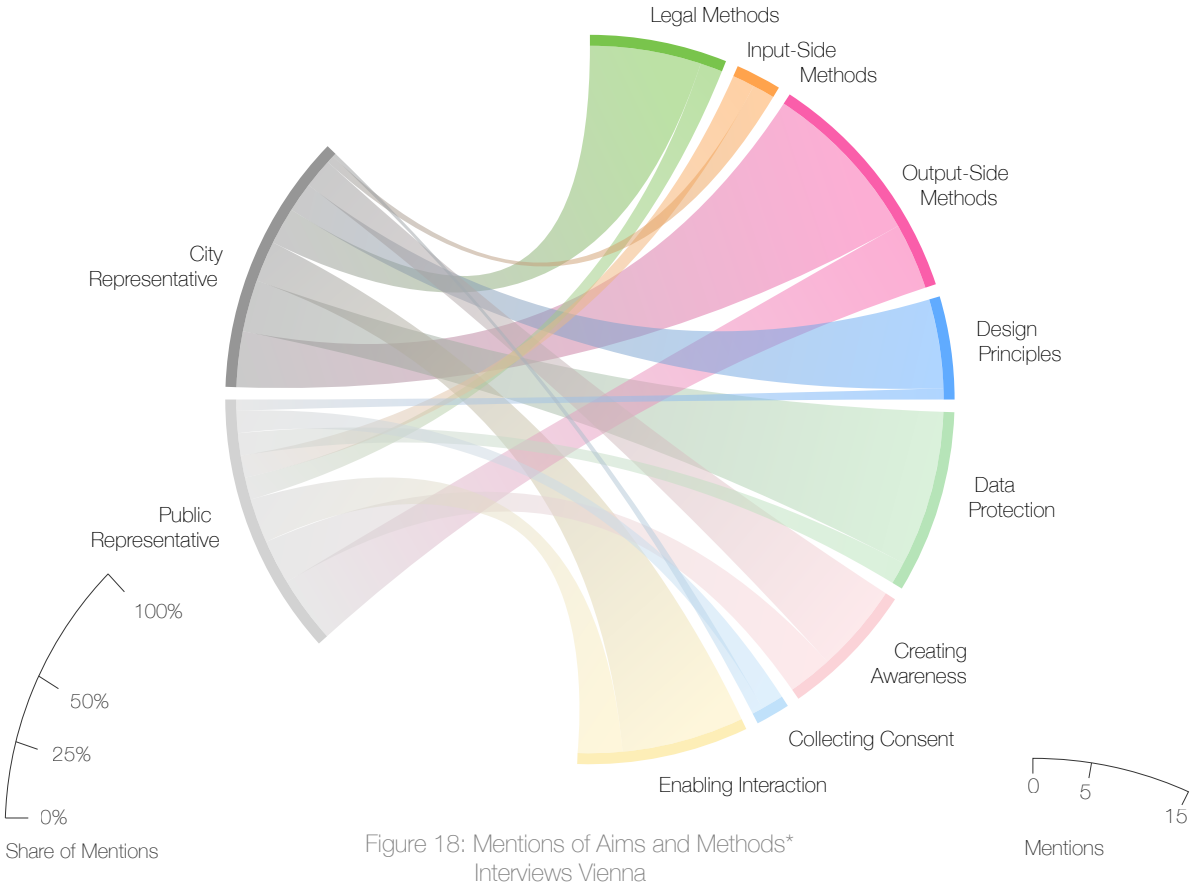


Figure 18: Mentions of Aims and Methods* Interviews Vienna

Similar to the other cities, in Vienna Output-Side Methods were mentioned the most by both city representatives and public representative. However, Legal Methods played a bigger role in Vienna than in the other cities. Correspondingly, the aim Data Protection was also mentioned most often (n=16) but played a more significant role for the city representatives than for the public representative where it accounts for approximately 20% of the mentions in total as opposed to approximately 10%.

[16] hackacity.eu last accessed 25.08.2022

Only the aim Collecting Consent and Input-Side Methods were more often addressed by the public representative than by the city representatives.

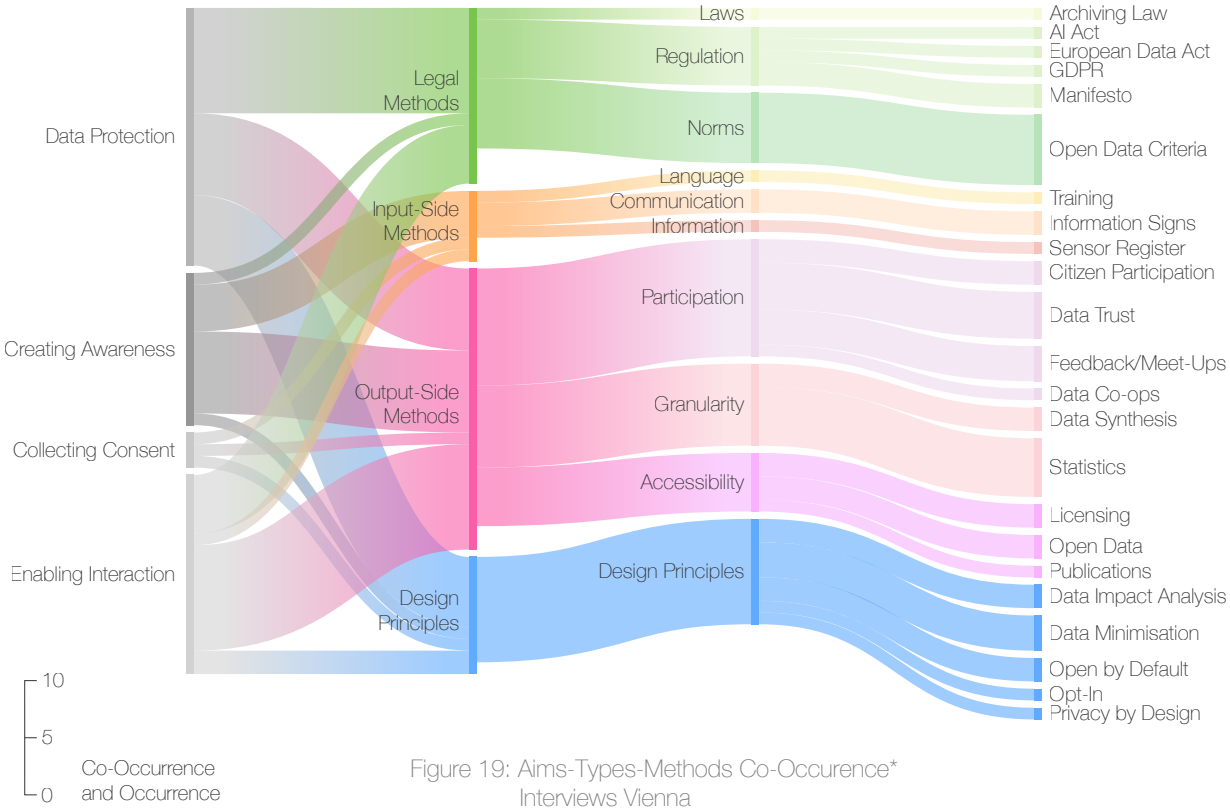


Figure 19: Aims-Types-Methods Co-Occurence*
Interviews Vienna

The high occurrence of the aim Data Protection co-occurs primarily with Legal Methods. However opposed to the other cities, Legal Methods have also been strongly discussed in the context of the aim Enabling Interaction. Likewise, Design Principles also play a stronger role regarding the aim Data Protection.

Among the Design Principles, the principle data minimisation, i.e the method to collect as little data as possible for a specific task, was discussed frequently (n=3) in Vienna. Similarly, a data impact analysis was also discussed. This is useful to delineate the use cases of data beyond the current objective and what impacts these use cases could have, i.e. if the data can be abused.

One norm that has been mentioned repeatedly is Criteria for Open Data. While usually the decision whether data should be published or not is made by the data controller (e.g. a specific city government body), this norm can help facilitate the decision making process and make as much data available as possible. Another specific method mentioned in Vienna are data trusts or data spaces, that means shared (digital) spaces multiple data consumers can access without making the data completely available. In this context, the European project Gaia-X was mentioned repeatedly that aims at building digital sovereignty by providing the technical infrastructure for such data spaces among other aims. Similar to the already occurred method of granularity changing in a dataset through statistics, the method of data synthesis in cooperation with the company Mostly AI[17] was discussed in Vienna. Here, a specific dataset containing personal data is altered by an algorithm that creates a faux-dataset with similar statistical values that can be published. The proposed European

[17] mostly.ai last accesssed 22.08.2022

AI Act and European Data Act were discussed as an upcoming method to limit what data can be used for algorithms and to facilitate data sharing.

5. Discussion of Findings

The scoping review and the expert interviews have been conducted to delineate the extent of the topic data as a commons in urban data governance on the one hand and to identify appropriate methods allowing an equitable governance of data on the other. The aim of these research methods is to answer the overarching research question what types of methods exist for an equitable governance of spatial behavioural data as a common good in European cities. For this, the scoping review has identified the disciplines active in the field as well as the topics they covered. Furthermore, the topics and how they are discussed specifically were examined and concerns for data as a commons identified. The interview analysis has identified methods in the practical discourse, shown the different understanding of data as a commons in cities, and uncovered the main difficulties, limitations and threats relevant for the governance of urban data.

As already discussed in the intermediate discussion, the scoping review has revealed a broad landscape of disciplines and topics they cover in the field of data as a commons. Specifically, concerns of data stewardship, ownership, digital rights, as well as questions of data availability and usages have been discussed. Furthermore, the discussion has shown that approaches of top-down management of these concerns have limited effects and awareness-raising and empowerment of the general public in the area of urban data is essential.

This finding is generally in line with the findings of the expert interviews. Here, the analysis has shown that the importance of equitable governance of data is a relevant concern that is however still an endeavour the cities want to reach (ref. fig.6: Different Understandings of Data as a Commons). For this the cities strive for finding ways to let their citizens participate, educate them, and give them agency in their cities' digitisation projects. Consequently, the prominence of Output-Side Methods in all cities, i.e. methods that address questions of data accessibility and participation, is unsurprising. Likewise, the prominence of the aims Creating Awareness and Enabling Interaction are a reflection of this endeavour. The limitations and difficulties the cities are facing in this pursuit stem primarily from Digital Literacy, i.e. technical capabilities of people both within the city governments and in the general public, and lack of appropriate Procedures alike (ref. fig.7: Main Limitations and Difficulties).

The problem of Invisibility of Benefits points in a similar direction of lack of awareness, thus rendering the importance of empowerment through giving technical knowledge and tools even more crucial. Beyond that the concern of missing Data Quality and the difficulty of (technical) Maintenance surrounding data point to reliability and costs generated by digitalisation project. Correspondingly, the perception of threats stated by the interview partners and what these difficulties and limitations could lead to lie in Too Much Trust in Data, i.e. reliance, and also the threat of Digital Divide, outpacing of some people because of lack of digital literacy (ref. fig.8: Tragedies of the Data Commons). The findings thus point to very clear and differentiated views on questions of data in cities from both city representatives' side and public representatives. Reflecting on the individual cities, the analyses has shown that compared to the other cities an elevated focus is given on methods of information and communication in Amsterdam (ref. fig.13). Aligning with the high focus on citizen empowerment the city of Barcelona is pursuing (Ajuntament Barcelona 2018; Calzada 2018) Output-Side Methods play an even more important role here than in all cities in general (ref. fig.15). In Porto, the methods mentioned appear to be equally comprehensive as in all the cities (ref.

figs.11;17). In Vienna, the focus on Regulation and Norms is elevated as opposed to the other cities (ref. fig.19).

Between the findings of the two methodological approaches scoping review and expert interview, similar topics and concerns and some particularities between the two emerged. Concerns over distribution of, accessibility to, and stewardship over data were the overarching topics in both parts that resulted in the need of empowerment and sensitisation for these topics in the general public. The topic Digital Rights are discussed primarily focussing on privacy and rights of participation in the scoping review. While the presence of the aims Data Protection and Creating Awareness in the interviews reflects these concerns, the question of Collecting Consent plays a bigger role in the interviews than in the identified literature of the scoping review. Here, the question of how to build consent focusses not only on legal consent but also on consent through stewardship. This suggests that if data subjects are empowered to decide what data about themselves is gathered and used for specific implementations the identification and acceptance of these implementations rises. Especially visible in the interview analysis by the co-occurrence of Output-Side Methods and the aim Collecting Consent. This supports the second hypothesis, the empowerment of citizens for building the understanding data as a commons. Further, the hypothesis is also reflected by the dominance of methods referring to accessibility and participation in the interviews as well as by the relevance of participatory topics in the review like Data Co-Operativism, Open Data or Data Trusts. This shows that building of awareness for data gathering and digital rights is not only a concern identified in the literature but also an existing endeavour pursued in the cities.

Concerning the topic Open Data and Data Economics, the assessment of the literature identified in the review has shown that both concerns of data quality and lacking disclosure of data by businesses holding data results in economic inefficiencies and insufficient usage of data for both economic and public interests. In combination with the potential tragedy of the data commons through Saturation identified in the interviews, more specifically the loss of economic value through meaninglessness because of substitutability of the dataset at hand, this points to the problem that businesses that hold data have an economic interest to artificially scarcen the resource. This means that businesses holding data have an interest to withhold this data from other data consumers or usages in an effort to make it more valuable for their usages. This means that they have an interest in not only limiting general access to the datasets they hold but also impeding the ability of other stakeholders to gather data in a similar dataset. Following that idea, a businesses collecting data from mobile phones, e.g. location in urban environments or proximity to shops, has an interest in making the sensors on those mobile phones inaccessible to other data gatherers or even the data subjects themselves to prevent the individual data subjects from gathering the same data in a collaborative way. For this example of data gathering through mobile phones, the questions of data authorship and ownership becomes strikingly clear as the sensor network, i.e. the mobile phones themselves, are legally owned by the individual data subjects whose data is being gathered. Thus this conflict of interests between private businesses, secondary data consumers, the data subjects, and the common benefit hints to both the condition of data production as well as the distribution of the generated dataset.

Similarly, the discussion of data ownership in the scoping review as a strawman argument, i.e. not pointing to the question of actual ownership of data but rather to the fair usage and access to it, and the higher relevance of accessibility to data rather than ownership in the interviews imply that the

discourse focusses more on the discourse of digital rights in cities rather than economic justices. This further reflects on the third hypothesis that the borders between enclosed data, open data, and data as a CPR are vague and fluent. These positions in the interviews and the discussion of the literature suggest that the precise definition of whether data in general is owned by a specific entity or controlled collectively as a CPR is secondary to the question of specific data usages and accessibilities. That means even if data is partially enclosed, it can still benefit the public interest if the usages as well accessibilities to this enclosure are collectively set. This is further reflected by the different understandings of data as a commons that were identified in the cities. Here, the focus is equally on data as a CPR and on Data For the Common Good (ref. fig.6) suggesting that both the concept of control and usage of the resource are relevant. The two understandings are simultaneously not mutually exclusive; data can be considered a CPR or not and be serving the common benefit or not at the same time. Correspondingly, the references identified in the scoping review focussing on the topic Data as a Commons neither primarily discuss the question of collective ownership. Rather, the references focus on data serving the common good, e.g. for questions of sustainability (Creutzig 2021) or questions of individual privacy (Wong, Henderson and Ball 2022).

Consequentially, the question of who oversees the implementation of methods of governance that ensure that data is employed to serve a common benefit appears. Here, the role of the city government is prominent in the interviews for example by making data available, incentivising data sharing by private businesses, or through financially supporting data co-operatives. The role of the city government is equally present in identified literature, for example by creating and operating data trusts (Huq 2021). This points to the new responsibilities city governments need to live up to addressing questions of digital equity, data distribution, and sovereignty. While the identified aims pursued are largely addressed by specific methods, some aims have still been reported as not yet satisfying. Evidently, this suggests that the pursuit is not yet sufficient and that the methods identified are not yet comprehensive in the pursuit of aims. This lack of methods in the pursuit of aims reflects the first hypothesis that the conditions of data production, rules of its use, and the power relations between the stakeholders are not yet well defined.

However, the diverse responses for a possible tragedy of the data commons, namely putting Too much Trust in Data, the Digital Divide, and the Saturation of the resource (ref. fig.8), but also the prominence and sensibility for the lack of Digital Literacy (ref. fig.7) indicate that there exists awareness and understanding of difficulties within the city governments and also from the outside view on the cities' responsibilities.

The study thus shows that a broad sensibility for questions equitable governance of data exist. Further it has identified topics relevant in the theoretical discourse as well as difficulties and specific methods addressing an equitable governance. However, the formulation of methods pursuing the aims considered relevant are not yet always considered to be sufficient for the pursued aim. Thus there exists still a discrepancy between the aims deemed relevant and the methods pursuing them. Finally, the findings of the scoping review and the expert interviews allow the answering of the research question by systematising the methods in a conceptual framework described in the next chapter.

6. Conceptual Framework

6.1 Proposed Framework Matrix

The scoping review and the interview examination have shown that for the theoretical and practical equitable governance of urban data, methods are being discussed and used that can be categorised by their type and by their aim. The different types identified are Legal Methods, Input-Side Methods, Output-Side Methods, and Design Principles. The aims mentioned in the interviews are protecting rights, Creating Awareness, Collecting Consent, and Enabling Interaction. While one method can pursue multiple aims, only few methods can be categorised between two types of methods. For example, only Data Co-Operatives could be understood as both input-side and output-side method.

Based on these findings about specific methods and their aims, a framework for classifying these methods is elaborated. Whereas the first draft of the framework only consisted of the different types, the proposed framework includes the aims of the methods and thus evolved into a matrix. This matrix (fig. 20) is depicted on the next page including the methods identified in this study. The matrix can be read as follows:

On the left hand the four types Legal, Input-side, Output-Side Methods and Design Principles are listed, they split into relevant categories in the second column. In the third column the individual methods are listed in accordance with their types. At the top, the four aims pursued by the methods are listed: Protecting Rights, Creating Awareness, Collecting Consent, and Enabling Interaction. In the resulting matrix between methods and aims, functioning for each method is described horizontally in description boxes. The outline of these boxes refers to the dimensions of equitable governance the methods address; a dotted line for the Conditions of Data Production, a dashed line for the Distribution of the Resource, and a closed line for the Use of the Resource.

For example, the already implemented method GDPR of the type Legal Methods pursues the aim Protecting Rights by limiting the gathering and transfer of personal data and requiring modes of consent, thus addressing the Conditions of Data Production. Additionally, the method also pursues the aim Enabling Interaction (with the data) by foreseeing the Right to Data Portability, i.e. addressing the Distribution of the Resource.

Further, the matrix also shows the interactions between the individual functions and methods with a dashed line and an arrow indicating the direction of interaction where applicable. In the example of the method GDPR, the Right to Data Portability is utilised by Data Cooperatives that individually claim data from data gatherers and aggregate it together.

The conceptual framework intents to give a basis to talk about methods of the shared resource data as a common good whether they are already implemented or not. Thus, the framework is not meant to be comprehensive. Rather it can help cities identify methods that have not been implemented yet and see what field or pursued they lack action in. Likewise, the framework can also be used to assess different methods and compare them to one another.

On the subsequent pages specific methods are closer described and their functioning explained, followed by examples of interactions between these methods.

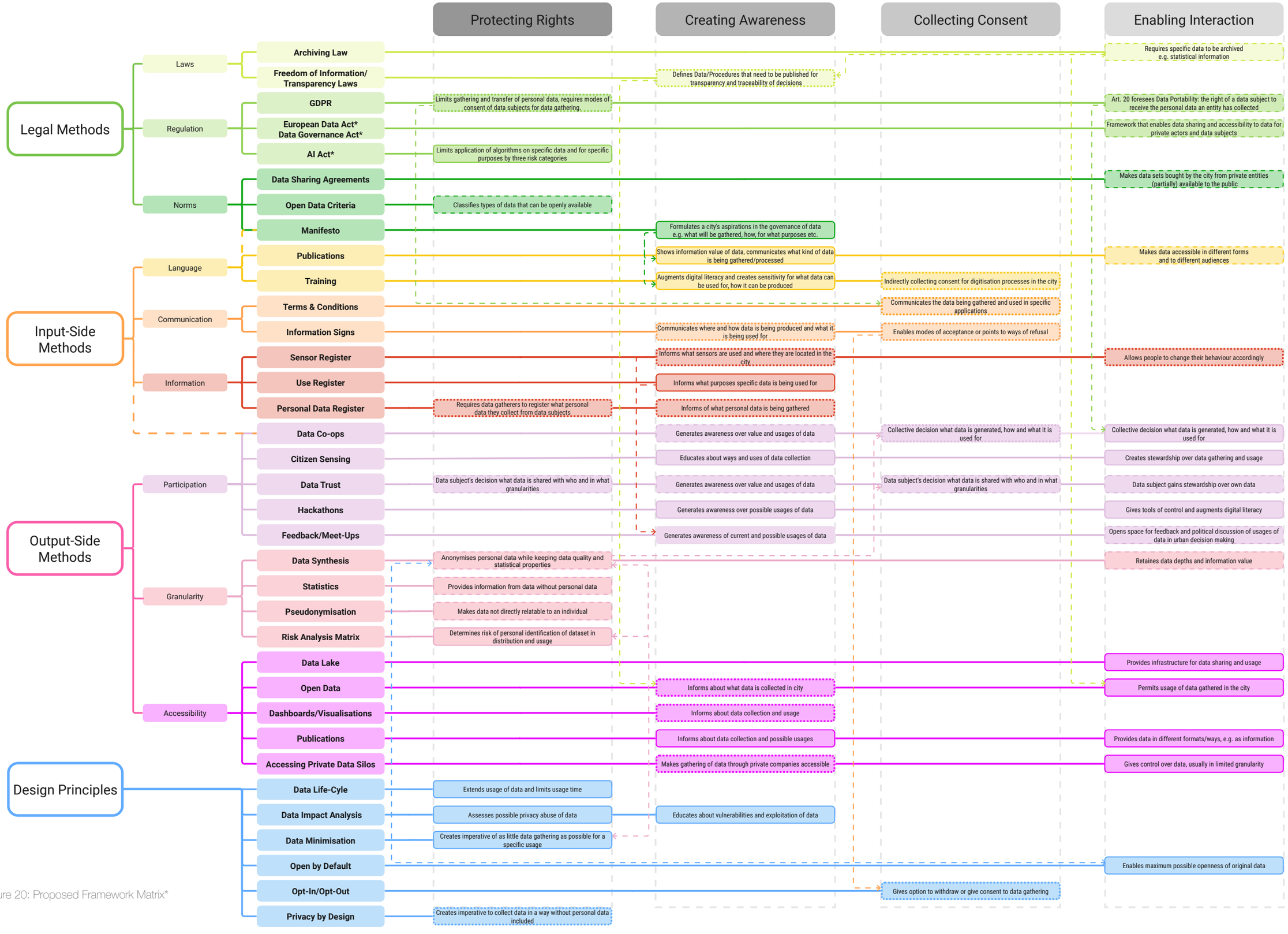
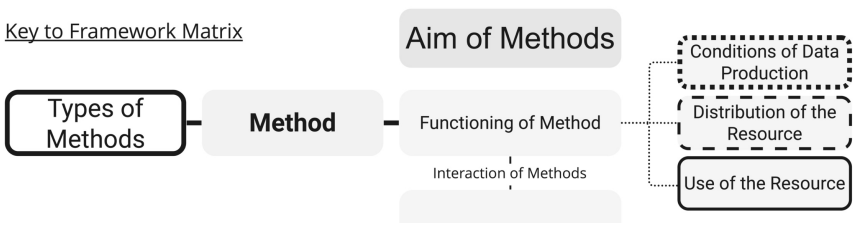


Figure 20: Proposed Framework Matrix*


Key to Framework Matrix
Aim of Methods
Types of Methods
Method
Functioning of Method
Interaction of Methods
Conditions of Data Production
Distribution of the Resource
Use of the Resource

6.2 Review of Specific Methods

Criteria for Open Data

An important element that is often reported in the interviews, is an overview of what kind of data should be made open and what cannot be made open. Having general criteria for what data can be published without concern, could help city servants in deciding and augment transparency. For this, the Open Data Institute (2020) has defined five categories of openness on an Open Data Spectrum depending on how closed or open the data is. From internal access, named access with specific contracts, via group-based access for data that can be accessed through authentication, public access with limited usage rights through licensing, to Open with licenses that do not restrict the usage. Additionally, they offer examples of what these data types can be for specific application areas like mobility, or telecommunication. However, this spectrum does not offer precise criteria for decision makers in city governments to decide what data can be shared without issues and what data must be limited.

Figure 21: AI Act, own representation based on (European Commission 2021: 12-14) and (Liguori 2022)

AI Act

The EU AI Act is a proposed regulation that aims at defining for what uses algorithms can be employed and what kind of data can be used by them. The uses are classified in four different categories with relevant limiations: no risk (product recommendations etc.; no limitation), limited risk (chatbots, image manipulation; user transparency), high risk (critical infrastructure, education, law enforcement; external and human supervision, relevant datasets), unacceptable risk (social scoring, remote biometric sensing; complete ban).

European Data Act, Data Governance Act

The two acts are proposed regulations. The Data Act aims at standardisation of data gathering to facilitate data sharing for the data subjects/generator. For example having a standardised format that the data subject can download from one provider to give it to another one (European Commission 2021). The Data Governance Act aims at facilitating data sharing between stakeholders. One of the proposed elements are data spaces (European Commission 2020: 10) that can be compared to data trusts. In such spaces, data producers and data consumers can come together and create a shared (virtual) space with shared data. For example, a city that has gathered data about citizen movement could make this data available to the public transport company in such a data space. With the proposed legislation the European Commission aims at creating a harmonised data economy, in which new business like data brokerage or data intermediaries create value by connecting data from individual data subjects, but also facilitating data altruism, i.e. sharing data without an economic benefit (European Commission 2020).

Sensor, AI and Use Register

A method already in use in the City of Amsterdam[18], sensor, ai or data use registers are public repositories in which sensors or algorithms are listed that are installed in public spaces. They can include publicly owned as well as private sensors or algorithms. Such registers cannot only include the position of the sensor, but could also include meta data like field of vision, gathering method or the usage of the gathered data.

The use of such registers for private sensors in Amsterdam is currently voluntary, mainly bound to city governments' commitments to be transparent about their registers. However, a legal reporting obligation could augment transparency and also access private sensors situated in public, like private surveillance cameras.

Information Signs

Already a common element in public spaces, information signs often inform about camera surveillance. Yet, few other sensors like noise, pollution or Wi-Fi sensors that also collect personal information are signposted like cameras. Only rarely do these signs inform about what the gathered data is used for and what kind of options the data subjects have. If this information is given, the results are usually signs with a lot of text that is not effortlessly and quickly comprehensible. Here, a standardised visual language could help communicating the gathering and use of data. In the visualisation, the signs communicate not only how data is being gathered, but also how the data is being processed and how it is being used. In the example (Fig. 22), different decision making processes based on data gathering are described; a camera that feeds an algorithm for automatic management of traffic lights, a Wi-Fi sensor, that counts the number of devices in an area to determine the crowdedness, or a noise sensor that automatically limits traffic speed if a certain value is passed. These can trigger questions like, what data does the camera specifically gather; does it analyse only cars and thus only improves traffic light management for the flow of cars, while cyclists and pedestrians might not be captured by the algorithm. It can provoke the question whether the process chosen has blind-spots or can create meaningful results. Traffic signs can thus create more awareness for questions of digitisation than just informing about data gathering.

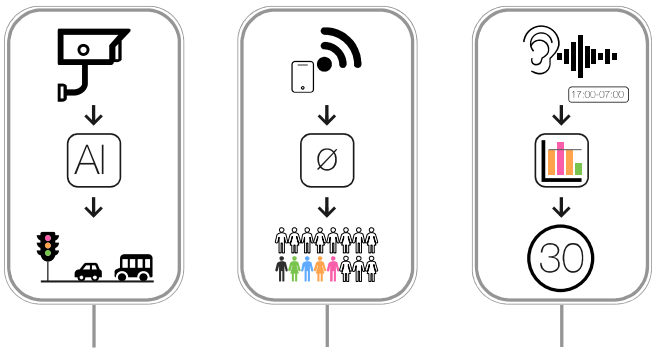


Figure 22: Ideation for Information Signs

Further, signs can also inform about modes of consent, opt-out options, or access and interaction to the gathered data. An example in Amsterdam, the aforementioned project *Shuttercam*[19], by the Responsible Sensing Lab goes further and experiments with letting the users shut off the sensor for a defined time. This could be used for sensors that do not need to gather data all the time.

[18] sensorenregister.amsterdam.nl last accessed: 22.08.2022
[19] responsiblesensinglab.org/projects/shuttercam last accessed: 22.08.2022

Statistics, Granularity

Statistical methods are a very common and proven method of data governance in cities. Through operations like averaging, creating a sum or calculating a relation between specific values in a dataset that contains personal data, data can be made accessible with a lower degree of granularity and thus not infringing personal data rights.

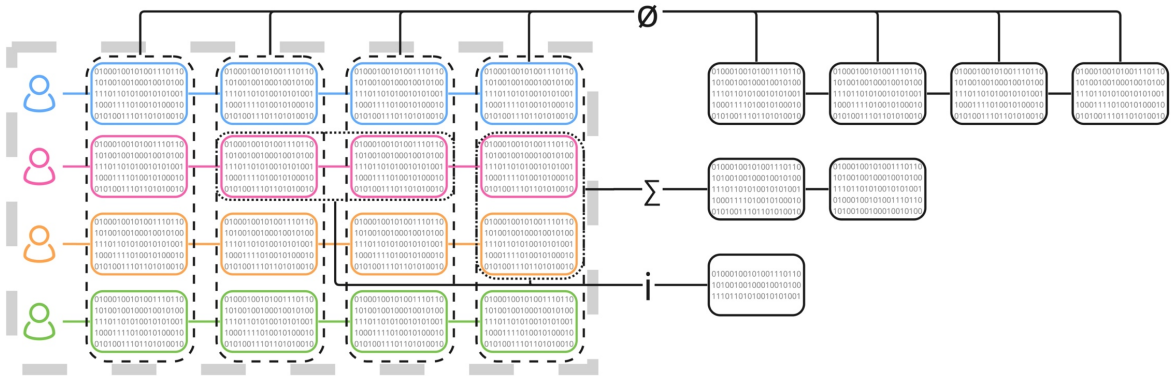


Figure 23: Statistical Methods applied to a Dataset*

A typical example of this is information about where people life in cities. The city gathers the data in the population register and thus know exactly where people life. The data is than aggregated for specific neighbourhoods, streets, or building blocks and made publicly available. The data consumer does not know where a specific person lives but knows how many people live in specific neighbourhoods.

Logically, the degree of granularity and thus the depth of information declines with statistical methods. This method does thus not allow for direct evaluation of the full depth of the data gathered.

Data Synthesis

This disadvantage of statistical methods can be substituted by data synthesis. In this method, a faux dataset is being generated based on a source dataset. In the visualisation, the source data contains personal data of four individual people. This can for example be a dataset with specific mobility data; chosen mode of transport, departure and destination point, route travelled etc. This dataset could not be made publicly available, as the individual could be easily identified. Therefore, the data remains unpublished (grey dotted line). Using an algorithm however, a faux data set is generated having similar to the same statistical values as the original data, while changing the individual data points beyond identification to the source dataset.

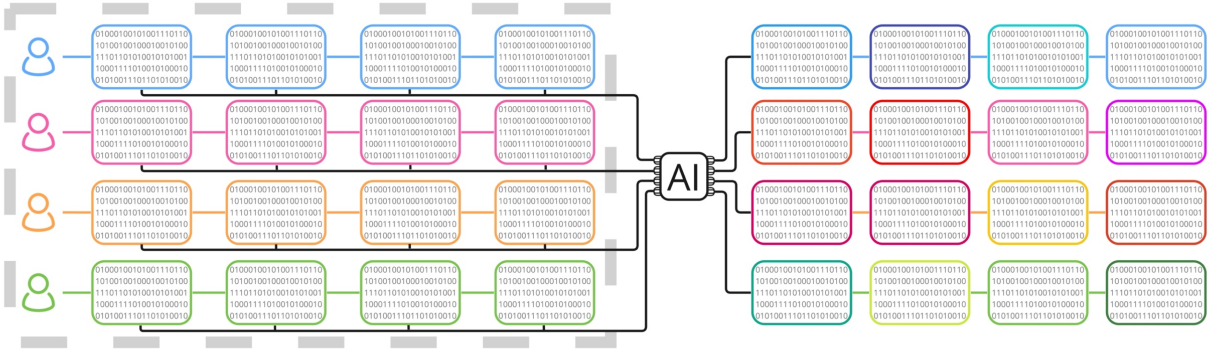


Figure 24: Conceptual Diagram of Data Synthesis*

This dataset can then be openly shared as identification of the data subjects is not possible anymore. The advantage this method has over statistical methods is that the dataset keeps a high degree of relevance. However, there is always some granularity lost as the fake data can never be as precise as the original data.

Data Trust

A data trust or data space is a virtual repository in which a data subject (e.g. a smartphone user, citizen) can share the data they have generated with defined data consumers under set conditions. In a data trust the data is supervised by a trustee (e.g. a governmental body, or independent instance), a data space is controlled by the data subjects or holders themselves. For example, a user that has generated movement data with their phone and wants to make this data accessible. They can decide which data consumers to share this data with. For example, giving full access to the data to the city government, selling the data to public transport companies with a lower degree of granularity, or giving a faux data set to scientist using the data synthesis method. Technically, the data space can be multiple spaces that share different information, or a technical process that uses statistical or data synthesis methods to provide data in different granularities to different data consumers. A data trust can also be maintained by a group of data producers or by multiple data consumers that use a certain data set in similar ways. For example, a city government and a public transport company using movement data; while the public transport company is interested in where most people come and go to, the city government is more interested in how they move.

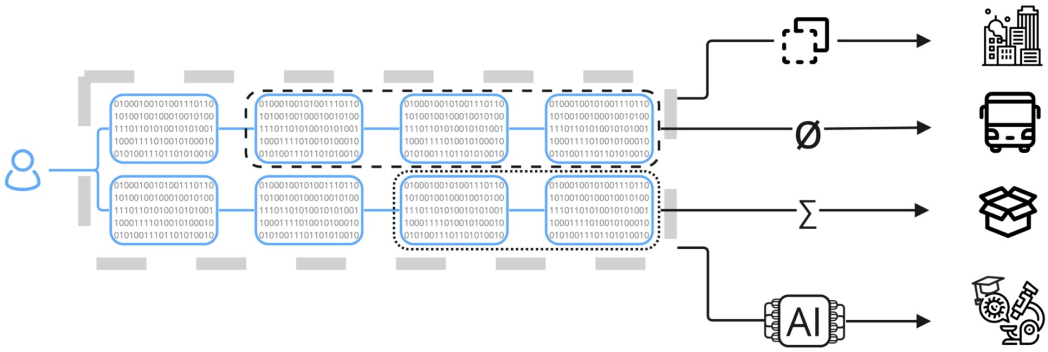


Figure 25: Conceptual Diagram of a Data Trust*

The idea of a data trust does not appear to be broadly in use yet. However, it seems to promise to address all four aims of the methods; it protects the individuals' rights by giving clear authority over the own data, and collects consent thereby, creates awareness by making the data usage and consumers visible, and enables interaction for the data generator by giving full access to their own data and access to the decision what it is being used for.

Data Lake and Open Data Portals

A Data Lake can be understood as a storage of multiple different data sets. The datasets can contain structured, semi-structured or unstructured data in different formats from diverse data sources, like movement data from public transport companies or individual movement data from private data subjects. The data lake can be openly accessed or have limited access to specific data consumers.

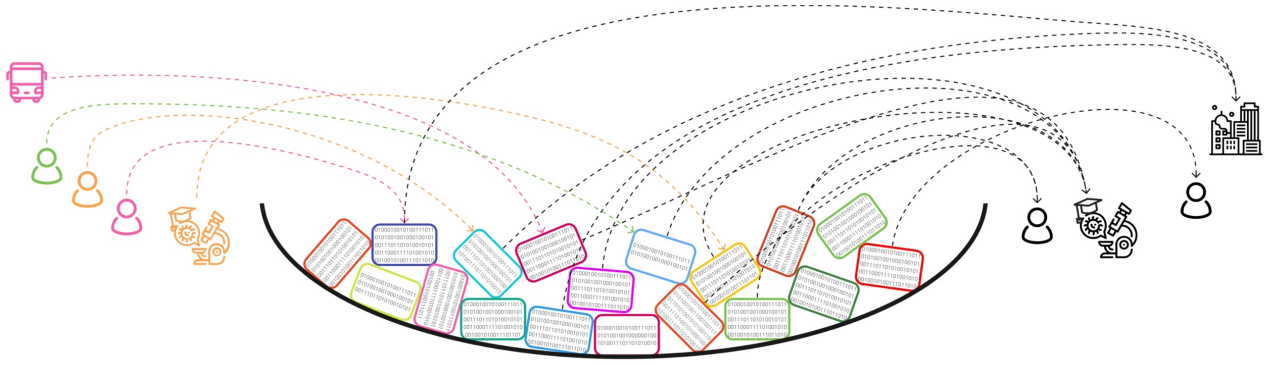


Figure 26: Conceptual Diagram of a Data Lake*

In an Open Data Portal in contrast, data is provided in a standardised, machine-readable format(s), and the datasets are kept up to date. The advantage of a data lake is that correlations between different data sets can emerge, and the data is provided in a use-agnostic way; without a specific use of the data in mind. However, as there is no assessment of data quality, relevance or amount of data, a data lake risks to become an excessive data storage without clear overview what data exists within.

Data Co-ops and Citizen Sensing

In a Data Cooperative, members of the cooperative gather their personal data from either the businesses that have collected from them or from their own sensor networks and aggregate them in a collected ecosystem to collectively govern and use this data set.

An example of this is the Catalan project Salus Coop[20] where members retrieve their personal data from their health insurance companies, aggregate all the individual datasets, and make them available for research.

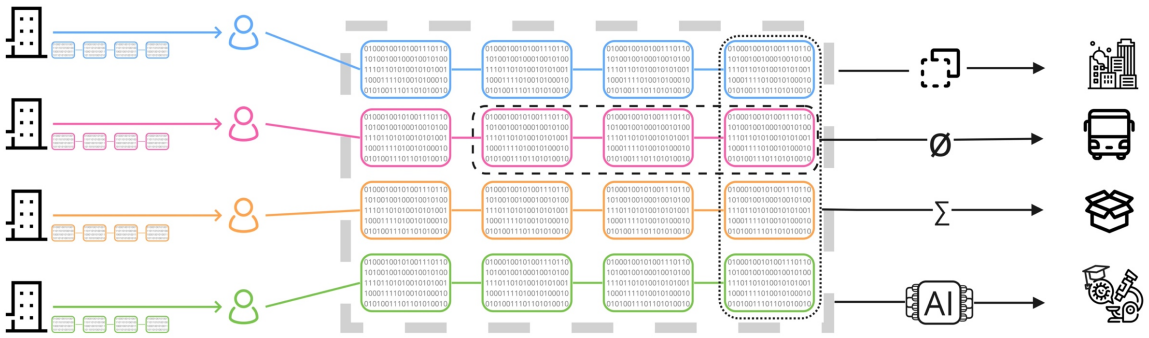


Figure 27: Conceptual Diagram of a Data Co-Operative

While a Data Co-Operative focusses on the joint gathering and distributing of the data, a citizen sensing project is about the joint generating of data sets. For example, a project of installing air quality

[20] saluscoop.org

sensors in a neighbourhood to address a specific issue. In such projects, the physical impact is in the foreground, while the data gathered is a means to the projects' end.

Data Life-Cycle

The method of a data life cycle is to assess the different uses specific data can be used for and to define when the data's life-cycle has finished; i.e. defining when it will be deleted. This can for example result in an automatic deletion of data after a certain time span or after a certain amount of uses of the resource and thus augment its value.
As an artificial shortening of the resource, limiting its usages would however diminish the advantage of data to not deteriorate through its usage as a non-rival good. However, an automatic deletion of specific data points could improve Data Protection and also the relevance of the data points. With the assessment of relevant usages, the method can also result in a change of how and what data is collected precisely in a project, being the case for example, when the change could allow more usages and relevancies. An example could be a traffic sensor not only measuring the number of vehicles passing by on a road but also the kind of vehicle, car, bike, or truck.

Data Impact Analysis

A Data Impact Analysis is a way of technology consequence assessment being applied to evaluate the risk deriving from specific data being gathered and what ways of misuse of the data could take place. For example, if a way of data gathering allows conclusions over marginalised groups and could thus be used to target or discriminate against these. In a way, it is the counterpart to the Data-Life-Cycle; instead of searching for positive secondary uses, the Data Impact Analysis searches for negative secondary uses. Furthermore, the analysis can also include questions of dependence on and vulnerabilities of data or technology. If for example the gathering of data draws a critical process dependent on it or could risk the safety of it, additional safety measures are necessary (e.g. enclosure of the data even though it does not include personal information).

Risk Analysis Matrix

A Risk Analysis Matrix addresses the risk of identification of individuals in a given dataset. For example, in a dataset with lower granularity that has been created from a dataset with personal information using statistical methods there might still exist a potential for identifying the individuals from the original data set. The risk analysis determines how high this risk is and formulates criteria of acceptance or rejection. Such an assessment can be done automatically by an algorithm calculating the risk and guarding access to data automatically.

Data Minimisation and Privacy by Design

Data minimisation is a term also promoted by the GDPR, it describes the principle to collect as little data as possible to fulfil a specific use. The aim is to avoid unnecessary gathering of personal data as well as overwhelming amounts of data. The method is thus possibly in a logical competition with a Data Life-Cycle.
Privacy by Design refers to the principle to collect data in a way in which no personal data is gathered. An example of this can be a traffic counting that can be done by using a camera that identifies

vehicles (thus gathering personal data) or through an electromagnetic sensor in the ground (thus not knowing which vehicles are driving above). The principle Privacy by Design would prioritise the latter method of data gathering.

Pseudonymisation and Anonymisation

Pseudonymisation refers to a de-identification of a dataset by replacing the unique personal identifiers (e.g. name, social security number etc.) with an alias. Thereby detaching a dataset from direct personal identification. However, a pseudonymised dataset can still uniquely identify the individual. Either if the pseudonym and the identifier of the dataset are revealed (i.e. leaked) or through the analysis of the data set, i.e. by comparing the dataset to the real world realities or connecting it to another data set. For example, if a pseudonymised dataset refers to a person that travels to a specific place daily the data can be easily re-assigned to the person by surveying this place and thus also identifying all the other data points about this person in the data set.
In contrast, an anonymisation refers to the removal of all personal identifiable information in a dataset. This can happen by completely deleting the relevant information thus disconnecting all individual data points from one another. That means for example in a dataset that has the information of one individual going somewhere and then going somewhere else, the two journeys would be completely detached from one another after an anonymisation. To ensure anonymisation however, an alteration of the data might even be necessary if there still exists risk of identification. For example, if through a specific behaviour individual data points can be connected to one another, and consequently to a person. Here, a Risk Analysis Matrix can for example assess and quantify this risk.
The term anonymisation is often falsely used when in fact referring to a pseudonymisation.

Accessing Private Data Silos

The problem of enclosures of data through private businesses that could be used for a public good can be addressed in two ways. Either by incentivising data disclosure through purchase or exchange of resources that are important to the businesses or through regulation. While buying data is more common and easier, some cities have started regulating digital platforms like Uber and Airbnb to be able to access their data.

6.3 Interaction of Methods

The assessment of the different methods has already suggested that the individual methods interact by encouraging or excluding one another. As discussed by Smichowski (2019) the power lies within the combination of individual governance methods, rather than the perfect elaboration of one specific governance model. The whole image of methods pursuing an equitable governance of data emerges when the methods are combined strategically and work in a same direction to pursue a certain goal. For example, the requirement of GDPR to collect consent for personal data gathering results in the method of putting this information in terms and conditions that the data subjects must agree to before personal data can be collected. The aim Enabling Interaction pursued by GDPR through the concept of Right to Data Portability is even more powerful. This right interacts with the method Data Co-Operatives. As described earlier, some data co-operatives have made use of this right to request their individual data from private businesses having collected their data. The data individually regained is than aggregated together to control it collectively.

A Risk Analysis Matrix can be used together with specific methods like a Data Synthesis or Statistical Methods and could require specific licenses for the use of data. For example, if the disclosure of a specific dataset is connected to a certain risk, an automatic requirement for synthetisation of the dataset or a license limiting the use could help reduce this risk.

Similarly, the principle Open by Default could be expanded to also include personal data if it interacts with the method Data Synthesis. Through an automatic anonymisation process of personal data, the full usefulness of data could be made publicly available.

However, there also exist methods that negatively affect or exclude each other. A Data-Life Cycle Analysis identifying multiple use cases for data for example could be in direct concurrence to the Design Principles Data Minimisation and Privacy by Design. If the analysis for example results in an extended use case by gathering more data than necessary for the primary use case this contradicts the principle to gather as little data as possible (Data Minimisation).

7. Conclusion and Reflexion

7.1 Conclusion

This study has explored the research question, what types of methods exist for an equitable governance of spatial behavioural data as a common good in European cities. To answer this question, a scoping review and expert interviews in four European Cities were conducted that enabled insights in both the theoretical extent of the topic as well as the practical issues of equitable data governance.
The scoping interview has revealed a broad spectrum of disciplines active in the discourse stemming primarily from law, urbanism, and environmental studies, over computer and data sciences to economics and political sciences. Topics covered by these disciplines focus primarily on questions of digital rights, data ownership, stewardship and civic empowerment but also on specific difficulties of data governance and economic inefficiencies in data economy.
The expert interviews on the other hand revealed that there exists broad sensibility for these issues and for the role of city governments in the empowerment of its citizens. Here, the study identified four aims pursued by methods of equitable data governance; Data Protection, Creating Awareness, Collecting Consent, and Enabling interaction. While a wide range of methods already implemented and planned were identified, not all the aims stated were already addressed satisfyingly for the interview partners. The methods identified can be classified into Legal Methods, Input-Side, and Output-Side Methods, as well as Design Principles.
The discrepancy between aims and methods addressing them supports the first hypotheses according to which the conditions of data production, the power relations between the stakeholders, city governments, citizens and private businesses are not yet well defined. Similarly, this is also supported by the dominance of discussed lack of data sharing in the identified literature. The strong focus on Output-Side Methods, i.e. methods addressing questions of participation and accessibility, in the interviews support the second hypothesis that in order to understand data as a common good, citizens need to be empowered. Likewise, the study has also found evidence supporting the third hypothesis addressing the vagueness of borders between enclosed data, open data and data as a CPR. However, with the advanced finding that actual ownership of the data is not the concern but rather the question of its public availability and usage for the common benefit.
The final outcome of the study in answering the research question consists of a framework for classifying methods of an equitable governance of data consisting of both the types of methods identified in the study and the aims they address. This proposed framework aspires to help cities find areas and methods of data governance they have not addressed yet on the one hand and on the other hand help to systematise the general discourse on methods of equitable data governance in cities.
The study has shown that digital rights are neither merely digital nor exclusively legal concerns. Rather, they need to be addressed inter-disciplinarily with solutions that stem from legal fields protecting personal data rights, technological disciplines enabling traceability of data gathered, all the way to planning fields considering the significance of data that shape the built environment and the people living within.

7.2 Reflexion

If four years ago when GDPR came into effect someone would have told me that it will affect the built environment, I would have looked at them in disbelief. This study has taught that through decisions in urban processes like mobility that are based on personal data gathered it does affect how a city evolves thus making the question of data stewardship relevant even for urban planners.

Both the scoping review and the expert interviews have provided significant and satisfactory results for the answering of the research question. However, to meet the time and character limitations of this study, the analysis of the literature identified in the scoping review focussed primarily on the general discussion of topics and methods they covered. It would be interesting to further investigate how these methods are discussed by identifying more specific literature focussing on methods of data governance. While this would be beyond the scope of this study, further research into differences between theoretical and practical application of governance methods would therefore be relevant for the ongoing development of the research field.

Similarly, as the study combined two cities that specifically voice a commons-based approach (Amsterdam and Barcelona) and two cities that rather voiced a citizen-centred approach (Porto and Vienna), but did not find a major difference in understanding of the resource in practical terms, it could be relevant to also look at the topic broader and not focus on cities that voice commons-based or citizen-centred approaches to see if differences in handling would emerge within the European context.

The study has identified the different disciplines active in the discourse and examined the way they discuss the topics relevant to them. Likewise, the interviews revealed a broad range of methods being used. However, there exists no assessment of the impacts and effects these methods have on citizen involvement and urban development in the long run. An exploration into the impacts of the different methods on participation and on the built environment and how these impacts are perceived differently from the individual disciplines, e.g. legal, urbanism, and data scientists, would therefore return interesting insights.

The study focussed on the theoretical discourse among scholars as well as the practical concerns among experts in the cities. However, the people affected by the discourse, the data subjects and citizens in the cities, and their perceptions were not examined. Here, further research into the implications of data as a commons for data subjects themselves would be a relevant path.

The proposed framework is a starting point that intents to help cities identify new methods and ease the discourse on methods of data governance. As it is not meant to be comprehensive, new methods could fit in easily and emerging aims and classes could be meaningfully incorporated.

8. References

Ajuntament Barcelona (2018): Barcelona City Council Digital Plan. Government Measure Concerning Ethical Management and Accountable Data: Barcelona Data Commons. Barcelona: Ajuntament de Barcelona.

Austin, Lisa and David Lie (2021): Data Trusts and the Governance of Smart Environments: Lessons from the Failure of Sidewalk Labs' Urban Data Trust. In: Surveillance & Society 19(2), 255-261.

Beckwith, Richard, John Sherry and David Prendergast (2019): Data Flow in the Smart City: Open Data Versus the Commons. In: Michiel de Lange and Martijn de Waal (eds.): The Hackable City. Singapore: Springer, 205-221.

Benfeldt, Olivia, John Stouby Persson and Sabine Madsen (2020): Data governance as a collective action problem. In: Information Systems Frontiers 22(2), 299-313.

Benjamin, Garfield (2020): From protecting to performing privacy. In: The Journal of Sociotechnical Critique 1(1), 1-30. doi: 10.25779/erx9-hf24.

Benkler, Yochai and Helen Nissenbaum (2006): Commons-based Peer Production and Virtue. In: Journal of Political Philosophy 14(4), 394-419. doi: 10.1111/j.1467-9760.2006.00235.x.

Bevir, Mark (2012): Governance: A Very Short Introduction. Oxford: Oxford University Press.

Bloom, Greg (2013): Towards a Community Data Commons. In: Brett Goldstein and Lauren Dyson (eds.): Beyond Transparency. San Francisco: Code for America Press, 255-270.

Butler, Judith (1988): Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory. In: Theatre Journal 40(4), 519-531. doi: 10.2307/3207893.

Calzada, Igor (2018): (Smart) Citizens from Data Providers to Decision-Makers? The Case Study of Barcelona. In: Sustainability 10(9), 3252. doi: 10.3390/su10093252.

Calzada, Igor (2020): Platform and data co-operatives amidst European pandemic citizenship. In: Sustainability 12(20), 1-22.

Calzada, Igor and Esteve Almirall (2020): Data ecosystems for protecting European citizens' digital rights. In: Transforming Government: People, Process and Policy 14(2), 133-147. doi: 10.1108/TG-03-2020-0047.

Calzada, Igor, Marc Pérez-Batlle and Joan Batlle-Montserrat (2021): People-Centered Smart Cities: An exploratory action research on the Cities' Coalition for Digital Rights. In: Journal of Urban Affairs online first, 1-26. doi: 10.1080/07352166.2021.1994861.

Campolargo, Margarida and Paulo Calçada (2018): Porto Integrated Action Plan. Porto: Câmara Municipal do Porto.

Choenni, Sunil, Mortaza S Bargh, Tony Busker and Niels Netten (2022): Data governance in smart cities: Challenges and solution directions. In: Journal of Smart Cities and Society online first, 31-51. doi: 10.3233/SCS-210119.

Chyi, Natalie and Yuliya Panfil (2020): A Commons Approach to Smart City Data Governance. https://www.newamerica.org/future-land-housing/reports/can-elinor-ostrom-make-cities-smarter/, Retrieved 10.02.2022.

Cities Coalition for Digital Rights (s.a.): About Cities Coalition for Digital Rights. https://citiesfordigitalrights.org/about-us, Retrieved 01.02.2022.

Concilio, Grazia and Francesco Molinari (2021): The unexploitable smartness of open data. In: Sustainability 13(15). doi: 10.3390/su13158239.

Creutzig, Felix (2021): From Smart City to digital urban commons: Institutional considerations for governing shared mobility data. In: Environmental Research: Infrastructure and Sustainability 1(2), 1-18.

Cuff, Dana, Mark Hansen and Jerry Kang (2008): Urban sensing: out of the woods. In: Commun. ACM 51(3), 24–33. doi: 10.1145/1325555.1325562.

D'Ignazio, Catherine and Lauren Klein (2020): The Numbers Don't Speak for Themselves. In: Catherine D'Ignazio and Lauren Klein (eds.): Data Feminism. Cambridge, MA: MIT Press, 149-172.

de Hoop, Evelien, Timothy Moss, Adrian Smith and Emanuel Löffler (2022): Knowing and governing smart cities: Four cases of citizen engagement with digital urbanism. In: Urban Governance online first. doi: 10.1016/j.ugj.2021.12.008.

de Lange, Michiel (2019): The Right to the Datafied City: Interfacing the Urban Data Commons. In: Paolo Cardullo, Cesare Di Feliciantonio and Rob Kitchin (eds.): The Right to the Smart City. Emerald Publishing Limited, 71-83.

Delcroix, Geoffrey (2017): Smart cities and innovative uses for personal data: scenarios for using data to restore the balance between public and private spheres. In: Field Actions Science Reports. The journal of field actions (17), 75-79.

Drexl, Josef (2021): Connected devices–An unfair competition law approach to data access rights of users. In: German Federal Ministry of Justice and Consumer Protection and Max Planck Institute for Innovation and Competition (eds.): Data Access, Consumer Interests and Public Welfare. Baden Baden: Nomos, 477-528.

European Commission (2020): Proposal for a Regulation of the European Parliament and of the Council on European Data Governance (Data Governance Act). Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN, Retrieved 20.08.2022.

European Commission (2021): Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artifical Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. Brussels: European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206, Retrieved 20.08.2022.

Fia, Tommaso (2021): An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons. In: Global Jurist 21(1), 181-210. doi: 10.1515/gj-2020-0034.

Fisher, Angelina and Thomas Streinz (2021): Confronting Data Inequality. In: Columbia Journal of Transnational Law 60(3), 829-956.

Foth, Marcus, Irina Anastasiu, Monique Mann and Peta Mitchell (2021): From automation to autonomy: Technological sovereignty for better data care in smart cities. In: Brydon T. Wang and C. M. Wang (eds.): Advances in 21st Century Human Settlements. Singapore: Springer, 319-343.

Gemeente Amsterdam (2019): Een Digitale Stad voor én van iedereen. Amsterdam: Gemeente Amsterdam.

Goodman, Ellen P. (2020): Smart City Ethics: The Challenge to Democratic Governance. In: Markus D. Dubber, Frank Pasquale and Sunit Das (eds.): Oxford Handbook of Ethics of AI. Oxford: Oxford University Press, 823–839.

Google (2022): Google Books Ngram Viewer. https://books.google.com/ngrams/, Retrieved 16.05.2022.

Greco, Gian Maria and Luciano Floridi (2004): The tragedy of the digital commons. In: Ethics and Information Technology 6(2), 73-81. doi: 10.1007/s10676-004-2895-2.

Hack, Yonne-Luca (2019): Liveable Smartness in Cities. A strategic Proposal for Architects to improve Liveability through Smart City Implementations. Master's Thesis, Chair of Architectural Informatics. Technical University of Munich.

Hardin, Garrett (1968): The Tragedy of the Commons. In: Science 162(3859), 1243-1248.

Helfrich, Silke (2014): Gemeingüter sind nicht, sie werden gemacht. In: Helfrich Silke and Stiftung Heinrich Böll (eds.): Commons: Für eine neue Politik jenseits von Markt und Staat (2. Auflage). Bielefeld: transcript Verlag, 85-91.

Hess, Charlotte and Elinor Ostrom (2007): Understanding Knowledge as Commons. From Theory to Practice. Cambridge: The MIT Press.

Hicks, Jacqueline (2022): The future of data ownership: An uncommon research agenda. In: Sociological Review (online first), 1-17. doi: 10.1177/00380261221088120.

Houser, Kimberly and John W Bagby (2022): The Data Trust Solution to Data Sharing Problems. In: Vanderbilt Journal of Entertainment & Technology Law, Forthcoming online first, 1-52.

Hummel, Patrik, Matthias Braun and Peter Dabrock (2021): Own data? Ethical reflections on data ownership. In: Philosophy & Technology 34(3), 545-572.

Huq, Aziz Z (2021): The Public Trust in Data. In: Georgetown Law Journal 110(online first), 1-58.

Iaione, Christian (2019): Legal Infrastructure and Urban Networks for Just and Democratic Smart Cities. In: Italian Journal of Public Law 11(2), 747-786.

Johnson, Jennifer, Anna Hevia, Rebecca Yergin, Shayan Karbassi, Adira Levine and Jorge Ortiz (2022): Data Governance Frameworks for Smart Cities: Key Considerations for Data Management and Use. In: Journal of Law and Mobility (online first), 1-25.

Jung Marques, Maria Angélica, Jamile Sabatini Marques, Blanca C. Garcia and Tatiana Tucunduva Philippi Cortese (2021): Contributions to Knowledge-Based Development through commons theory, using data as a common good. In: Journal of Environmental Mangagement & Sustainability 10(1), 1-25. doi: 10.5585/geas.v10i1.18231.

Kitchin, Rob (2014): The real-time city? Big data and smart urbanism. In: GeoJournal 79(1), 1-14. doi: 10.1007/s10708-013-9516-8.

Kitchin, Rob (2016): The ethics of smart cities and urban science. In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374(2083), 1-15. doi: doi:10.1098/rsta.2016.0115.

Labaeye, Adrien (2019a): The Role of Digital Commons in a Socio-Ecological Transition of Cities. Dissertation, Mathematisch-Naturwissenschaftliche Fakultät Humboldt-Universität zu Berlin.

Labaeye, Adrien (2019b): Sharing Cities and Commoning: An Alternative Narrative for Just and Sustainable Cities. In: Sustainability 11(16), 4358.

Le Breton, Marie-Anaïs, Hélène Bailleul, Jean-Baptiste Le Corf and Boris Mericksay (2022): La gouvernance des données urbaines entre territoire de projets et projet de territoire L'exemple de Rennes Métropole. In: Flux 127(1), 65-84. doi: 10.3917/flux1.127.0065.

Leclercq, Els M. and Emiel A. Rijshouwer (2022): Enabling citizens' Right to the Smart City through the co-creation of digital platforms. In: Urban Transformations 4(1), 1-19.

Lee, Jung Wan (2021): The Data Sharing Economy and Open Governance of Big Data as Public Good. In: The Journal of Asian Finance, Economics and Business 8(11), 87-96.

Lehtiniemi, Tuukka and Minna Ruckenstein (2018): The social imaginaries of data activism. In: Big Data & Society 6(1), 1-12.

Leventhal, Gerald S. (1980): What Should Be Done with Equity Theory? In: Kenneth J. Gergen, Martin S. Greenberg and Richard H. Willis (eds.): Social Exchange. Advances in Theory and Research. News York Plenum Press, 27-55.

Liguori, Giuliano (2022): AI Act: A Risk-Based Policy Approach for Excellence and Trust in AI. In: Medium. https://medium.com/codex/ai-act-a-risk-based-policy-approach-for-excellence-and-trust-in-ai-d29ce0d54e2, Retrieved 20.07.2022.

Lomotey, Richard K, Sandra Kumi and Ralph Deters (2022): Data Trusts as a Service: Providing a platform for multi-party data sharing. In: International Journal of Information Management Data Insights 2(1), 1-20.

Lupi, Lucia (2019): City data plan: The conceptualisation of a policy instrument for data governance in smart cities. In: Urban Science 3(91), 1-35.

Lutz, Brigitte (2019): Data Excellence. IKT-Strategie. Wien: Stadt Wien.

Madison, Michael J. (2020): Tools for Data Governance. In: Technology and Regulation 29, 29-43.

Mayring, Philipp (2015): Qualitative Inhaltsanalyse. Grundlage und Techniken. Weinheim: Beltz.

Mercille, Julien (2021): Inclusive smart cities: Beyond voluntary corporate data sharing. In: Sustainability 13(15), 1-13. doi: 10.3390/su13158135.

Micheli, Marina, Marisa Ponti, Max Craglia and Anna Berti Suman (2020): Emerging models of data governance in the age of datafication. In: Big Data & Society 7(2), 1-15. doi: 10.1177/2053951720948087.

Morell, Mayo Fuster, Anna Cigarini and Enric Senabre Hidalgo (2021): A framework for assessing the commons qualities of citizen science: comparative analysis of five digital platforms (online first), 1-21. doi: 10.31235/osf.io/pv78g.

Niaros, Vasilis (2016): Introducing a Taxonomy of the "Smart City": Towards a Commons-Oriented Approach? In: Triple C 14(1), 51-61.

Nikander, Pekka, Ville Eloranta, Kimmo Karhu and Kari Hiekkanen (eds.) (2020): Digitalisation, anti-rival compensation and governance: Need for experiments. Proceedings, Nordic Workshop on Digital Foundations of Business, Operations, and Strategy, Espoo, Finland.

Open Data Institute (2020): The Data Spectrum. https://theodi.org/about-the-odi/the-data-spectrum/, Retrieved 22.07.2022.

Ostrom, Elinor (1990): Governing the Commons. The Evolution of Institutions for collective action. Cambridge: Cambridge University Press.

Paprica, P. Alison, Eric Sutherland, Andrea Smith, Michael Brudno, Rosario G. Cartagena, Monique Crichlow, Brian K. Courtney, Chris Loken, Kimberlyn M. McGrail and Alex Ryan (2020): Essential requirements for establishing and operating data trusts: practical guidance co-

developed by representatives from fifteen Canadian organizations and initiatives. In: International Journal of Population Data Science 5(1).

Park, Kyung S. (2021): Data as Public Goods or Private Properties?: A Way Out of Conflict Between Data Protection and Free Speech. In: UC Irvine Journal of International, Transnational, and Comparative Law 6, 77-102.

Pierri, Paola and Elizabeth Wiltshire (2021): Cities and Citizens for Digital Rights. The Role for Public Participation in Digital Governance. In: PND 2, 29-44.

Prainsack, Barbara (2019): Logged out: Ownership, exclusion and public value in the digital data and information commons. In: Big Data & Society 6(1), 1-15.

Scassa, Teresa (2020): Designing Data Governance for Data Sharing: Lessons from Sidewalk Toronto. In: Technology and Regulation (online first), 44-56.

Shaffer, Gwen (2021): Applying a contextual integrity framework to privacy policies for smart technologies. In: Journal of Information Policy 11(1), 222-265.

Singh, Parminder Jeet and Jai Vipra (2019): Economic rights over data: A framework for community data ownership. In: Development 62(1), 53-57.

Smichowski, Bruno Carballa (2019): Alternative data governance models: Moving beyond one-size-fits-all solutions. In: Intereconomics 54(4), 222-227.

Stadt Wien (2019): Digitale Agenda Wien 2025. Wien: Stadt Wien.

Taylor, Linnet (2016): The ethics of big data as a public good: which public? Whose good? In: Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374(2083), 1-13. doi: 10.1098/rsta.2016.0126.

Tricco, Andrea C., Erin Lillie, Wasifa Zarin, Kelly K. O'Brien, Heather L. Colquhoun, Danielle Levac, David Moher, Micah D. J. Peters, Tany Horsley, Laura Weeks, Susanne Hempel, Elie A. Akl, Christine Chang, Jessie McGowan, Lesley Stewart, Lisa Hartling, Adrian Aldcroft, Michael G. Wilson, Chantelle Garritty, Simon Lewin, Christina M. Godfrey, Marilyn T. Macdonald, Etienne V. Langlois, Karla Soares-Weiser, Jo Moriarty, Tammy Clifford, Özge Tunçalp and Sharon E. Straus (2018): PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. In: Annals of Internal Medicine 169(7), 467-473. doi: 10.7326/m18-0850.

van Loon, Nathalie and Rosalie Snijders (eds.) (2021): Data commons in smart mobility – the road ahead? Proceedings, Information Society, Llubljana.

van Zoonen, Liesbet (2020): Data governance and citizen participation in the digital welfare state. In: Data & Policy 2, 1-17.

von Grafenstein, Max, Alina Wernick and Christopher Olk (2019): Data Governance: Enhancing Innovation and Protecting Against Its Risks. In: Intereconomics 54(4), 228-232.

Walravens, Nils, Jonas Breuer and Pieter Ballon (2014): Open Data as a Catalyst for the Smart City as a Local Innovation Platform. In: Communications & Strategies (96), 15-33.

Wernick, Alina, Christopher Olk and Max von Grafenstein (2020): Defining Data Intermediaries: A Clearer View Through the Lens of Intellectual Property Governance. In: Technology and Regulation 2020, 65-77.

Wong, Janis, Tristan Henderson and Kirstie Ball (2022): Data protection for the common good: Developing a framework for a data protection-focused data commons. In: Data & Policy 4, 1-31.

Zuboff, Shoshana (2019): The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power. London: Profile Books.

Zygmuntowski, Jan J., Laura Zoboli and Paul F. Nemitz (2021): Embedding european values in data governance: A case for public data commons. In: Internet Policy Review 10(3), 1-29. doi: 10.14763/2021.3.1572.

9. Annex

Data availability statement: The data generated in this thesis is only partially available. The interview transcripts are only made available to the supervisors of this thesis. The data for the visualisations is available upon request.

*Visualisations with an asterix are available as an interactive visualisation by clicking on the visualisation or following the link in the table of tables and figures.

9.1 Table of Tables and Figures

The links documented in the tables lead to interactive versions of the visualisations. The links unfortunately might expire a few months after the final hand-in of the thesis.

9.2 Interview Outline

The interview outlines appear on the subsequent pages, first in English and then in German.

Governing the Urban Data Commons – Master Thesis Caspar Kleiner

General Outline for an Open Interview

Date:
Interview Partner:

Opening

	Thank you + personal introduction
Introduction	Short Introduction in Research Topic and Question - Data Commons as a counter-narrative to Surveillance Capitalism, exploitation of the good data - Urbanism perspective: control over how city evolves goes to companies, top-down rather than through political discussion. - Technology & path dependency, biases, and the normative force of the status quo. Questions of stewardship and authorship - What does it mean for the resource data to be a common good, how does it become this? - Methods of governance between the different stakeholder; citizens, city, companies. - Three Dimensions of Governance What types of methods exist for an equitable governance of spatial behavioural data as a common good in European Cities? Aim: Outline of a framework for methods of an equitable governance of the good data.
Interviewee	Short outline of connection with the topic of interviewee If necessary, open questions
	Permission for Recording

Conduction of the Interview

3 Parts: Data as a Commons
Dimensions of Governance; methods, formats etc.
Temporal Dimension and Prioritisation
- Depending on interviewee -

1. General Discussion: Data as a Commons

Description	Commoning; the process of making a resource common
	How can data become or be considered a common good? In which instances and under which circumstances is this possible? What types of data can be considered a common good? Transaction, movement, meta, personal etc. What does an overuse of the resource mean in the area of digital and data commons? (Tragedy of the commons)

2. Three Dimensions of Governance

Primary question: What are methods of an equitable governance?
How can citizens participate and shape this process?
How do we create the prospect for citizens to participate?

Description	Three Dimensions of Governance: The Conditions of Data Productions (how/who?) The Distribution of the Resource (who/whom?) The Use of the Resource (What for?) For these dimensions methods are necessary that allow the process of negotiation of power balances between the different stakeholders. What methods are already implemented, experiences, effects, and effect mechanisms (i.e hackathons, open data portals, communication/ personal, prof. experience, best practices) Focus on spatial behavioural / movement data (!)
Questions	The Conditions of Data Production (how?) - What types of data is collected in the city today? - How and who collects this data?

	- What information and communication exists for the production/use? - Are there ways agree/disagree? (Opt-in/Opt-out etc.) - Design Principles; privacy by design Data Distribution (who/whom?) - How and whom is the data made accessible to? - Ways of agree-/disagreement? - How is the distribution communicated? - Does the provisioning of the resource create an identification of citizens with the gathering? - Focus Open Data o What data is made available? o When/how/by whom? o What data is not made available? Why? o What obstacles/preconceptions exist? o What stakeholders access the data? Use of Data (What for?) - What is the data used for? (examples of different stakeholders) - Ways of limitation/disagreement? - Evaluation of Cost-Benefit? - Focus Open Data: o Use of Licensing? o What stakeholders use the data? o Is it possible to trace the use? o Evaluation of use intensity Eventually, referral to interesting projects/applications
Synthesis	Short presentation of the drafted framework for methods, remarks

Prospect: What methods will be necessary to consider the resource data a common good?

Description	Also, Three Dimensions of Governance: Conditions of Production, Distribution, Use
Questions	Conditions of Production - What methods can help to create identification with urban data production? - How can a societal discussion about its conditions be made possible? - How can data gathering and the people who generated the data be brought together? Distribution - What other kinds of data could be made publicly accessible? - How can the distribution be traced, understood? - How can the distribution be limited by the people who generated the data? Use - How can the use be traced and controlled? - How could a limitation of the exploitation of the good be managed and controlled?

3. Temporal Dimension and Prioritisation of Methods

Description	Commoning as a Process: Commons are not, they are being made -> necessary momentum for political discussion and negotiation of conditions
Questions	- What methods can sensitize the public for the topics data sovereignty, and data production in urban contexts? - How can methods help create the connection between data gathering and urban development? - How do individual methods work in dependence of one another? - Possibly, chronology of methods? - Relevant network of stakeholders in city, civil society, and private companies - How can this process be communicated? General development of digitation and data in the individual cities - Societal discussion around topic of data sovereignty - Since when does the Open Data Portal exist? o Circumstances and background o Network of stakeholders - Smart City projects and understanding of term “Smart City“

<u>End</u>

	Thank you + open questions of interviewee + feedback
	Possibly, snowballing, interview + literature suggestions
	Follow up with research findings and presentation Thank you

Governing the Urban Data Commons – Master Thesis Caspar Kleiner

Offener Interviewleitfaden

Datum:
Interview PartnerIn:

Beginn

	Danke + persönliche Vorstellung
Vorstellung	Kurze Einführung in die Forschungsthematik und Frage - Data Commons als Gegenmodell zu Überwachungskapitalismus und Ausbeutung der Ressource Daten - Städtebauliche Perspektive; Einfluss über Entwicklung bei privaten Unternehmen; Top-Down anstatt politischer Diskussion - Technologie und Pfadabhängigkeit; Bias und normativer Charakter des Status-Quos. Fragen der Daten Souveränität und AutorInnenschaft - Was bedeutet es, dass eine Ressource wie Daten Allgemeingut werden; Methoden zwischen den verschiedenen StakeholderInnen; Stadt, Unternehmen, BürgerInnen (Formate etc.) - 3 Dimensionen der Governance What types of methods exist for an equitable governance of spatial behavioural data as a common good in European Cities? Ziel: Klassifikationsmodell für Methoden der gleichberechtigten Governance des Guts Daten
InterviewpartnerIn	Kurze Vorstellung, warum interessant für das Interview Ggf. Rückfragen desder InterviewpartnerIn
	Anfrage Genehmigung zur Aufnahme

Durchführung des Interviews

3 Teile: _ Allgemein: Daten als Allgemeingut
_ 3 Dimensionen der Governance; Methoden, Formate etc.
_ Zeitliche Dimension und Priorisierung
(Fokussierung je nach InterviewpartnerIn)

4. Allgemeine Diskussion: Daten als Allgemeingut

Beschreibung	Commoning; Der Prozess eine Ressource zum Allgemeingut zu machen
Fragen	Wie kann die Ressource Daten als Allgemeingut verstanden werden? In welchen Fällen, Gegebenheiten ist das möglich? Welche Arten von Daten können als Allgemeingut verstanden werden? Transaktion, Bewegung, Meta, persönliche Inwiefern besteht Gefahr zur Übernutzung der Ressource Daten (tragedy of the commons)?

5. Drei Dimensionen der Governance

Übergeordnete Fragen: Was sind Methoden einer Governance?
Wie können BürgerInnen an dem Prozess teilhaben, diesen gestalten?

Beschreibung	Drei Dimensionen der Governance Die Bedingungen der Erhebung (Wie/wer?) Die Distribution der Ressource (Wer/wem? Die Nutzung der Ressource (Wozu?) Für diese Dimensionen sind Methoden/ Maßnahmen nötig, die die Aushandlung der Machtverhältnisse zwischen den AkteurInnen erlauben.

	Welche Methoden sind bereits implementiert, Erfahrungen, Wirkungen und Wirkungsweisen. Klärung des Methodenbegriffes. Bezug zu räumlichen Bewegungsdaten (!)
Fragen	Bedingungen der Erhebung - Welche Arten von Daten werden in der Stadt erhoben? - Wie und von wem werden diese Daten erhoben? - Welche Hinweise und welche Kommunikation besteht über die Erhebung(/Verwendung, ? - Welche Möglichkeiten des Einverständnisses gibt es (Opt-In etc.)? - Design Prinzipien; Privacy by Design? Daten Distribution - Wie und wem werden die Daten verfügbar gemacht? - Gibt es Möglichkeiten des Einverständnisses /Beschränkungen? - Wie wird die Distribution kommuniziert? - Stellt die Verfügbarmachung eine Identifikation mit der Erhebung dar? - Fokus Open Data: o Welche Daten werden verfügbar gemacht? o Wann/wie/durch wen? o Welche Daten werden nicht verfügbar gemacht? o Welche Hürden/Einschränkungen gibt es bei der Bereitstellung? o Welche AkteurInnen greifen auf die Daten zu? Nutzung der Ressource - Wofür werden die Daten genutzt? - Möglichkeit zur Einschränkung der Nutzung durch BürgerInnen? - Gibt es eine Kosten-Nutzen-Evaluation? - Fokus Open Data: o Besteht Lizenzzwang? o Welche AkteurInnen nutzen die Daten? o Gibt es eine Möglichkeit die Nutzung nachzuvollziehen? Verweis auf interessante Vorzeigeprojekte, Anwendungen
Synthese	Kurze Präsentation des Entwurfs für das Klassifikationsmodell Ggf. Anmerkungen

Ausblick: welche Methoden sind denkbar/ notwendig, um das Gut Daten als Allgemeingut zu verstehen?

Beschreibung	Ebenfalls innerhalb der drei Dimensionen: Erhebungsbedingungen, Distribution, Nutzung
Fragen	Erhebungsbedingungen - Welche Methoden können helfen, die Identifikation mit urbaner Datenproduktion zu erhöhen? - Wie kann die Diskussion der Rahmenbedingungen ermöglicht werden?

	- Wie können Erhebung und Erhobene zusammengeführt werden? Distribution - Welche Art von Daten könnten allgemein zugänglich sein? - Wie kann die Distribution nachvollzogen werden? - Wie können Einschränkungen der Distribution aus Erhobenenseite aussehen? Nutzung - Welche Möglichkeiten der Nachvollziehbarkeit der Verwendung könnte es geben? - Wie könnte eine Einschränkung der Verwendungsmöglichkeiten funktionieren und durchgesetzt werden?

6. Zeitliche Dimension und Prioritätensetzung der Methoden

Beschreibung	Commoning als Prozess, nicht an einem Tag Privatgut und am nächsten Allgemeingut -> notwendiges Momentum für die politische Aushandlung der Rahmenbedingungen
Fragen	- Welche Methoden sensibilisieren für das Thema Datensouveränität und Erhebung im urbanen Kontext? - Wie können Methoden eine Identifikation mit der Stadtentwicklung schaffen? - Wie bauen Methoden aufeinander auf? - Ggf. zeitliche Folge der Methoden? - Relevante Akteurskonstellationen - Wie kann der Prozess kommuniziert werden? Entwicklung im Umgang mit Digitalisierung und Daten in den jeweiligen Städten - Gesellschaftliche Diskussion über das Thema Datensouveränität etc. - Seit wann gibt es das Open Data Portal? o Umstände und Hintergründe o Akteurskonstellation - Smart City Projekte und Verständnis von Smart City

Ende

	Danke, ggf. offene Fragen desder InterviewpartnerIn, Feedback
	Ggf. Snowballing, Interview- + Literatur Empfehlungen
	Danke, Auskunft zu weiterem Weg, Veröffentlichung und Präsentation

Governing the Urban Data Commons
Master Thesis at the Chair of Urban Development
TUM School of Engineering and Design
Technical University of Munich

Caspar Florens Kleiner
c.kleiner@tum.de
03673815