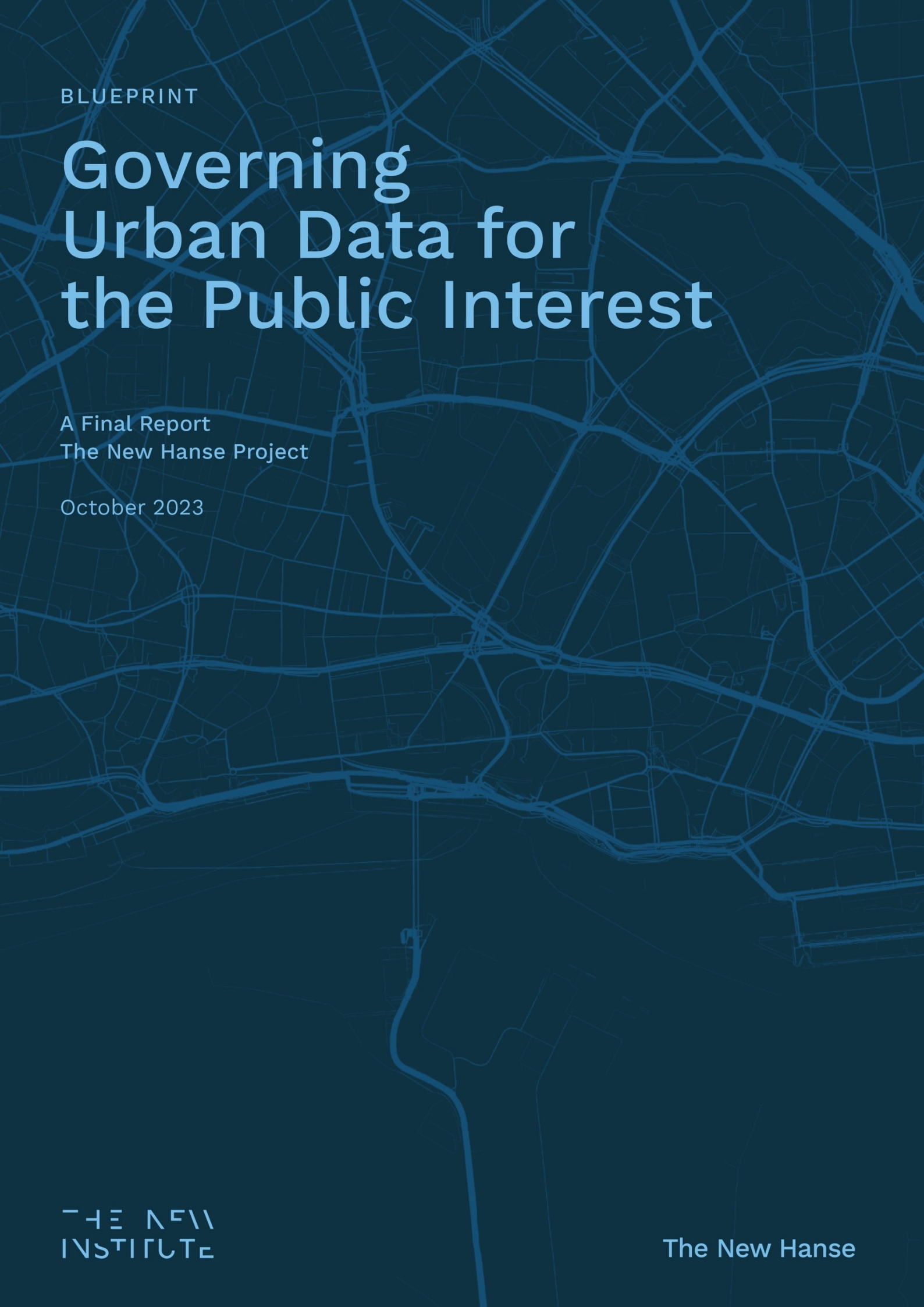
BLUEPRINT

Governing Urban Data for the Public Interest

A Final Report
The New Hanse Project

October 2023



THE NEW INSTITUTE

The New Hanse

Executive Summary

Cities are laboratories for democratic and sustainable innovation; they wield normative and regulatory power coupled with infrastructural capacity and close connection to citizens. They govern and shape the urban spaces in which citizens meet and interact. This urban space is becoming digital, which creates a new form of infrastructure: urban data understood as data collected in the public space or generated in the context of city procurement or financing. Urban data is essential to understanding and shaping how citizens can make use of the public space and how governments can make decisions and act in the public interest: Which needs are currently unmet by public transport? What do we need for a more effective mobility transformation? Where do we need more spaces for local communities and vulnerable groups? How can we provide better and more innovative public services?

Cities have only begun to tap into the potential of urban data. Although they make some of the data they collect available as open data to the public, most urban data is controlled by private companies that operate in the urban space and are reluctant to share this data. Yet, if urban data is understood as a common, cities should be able to access it to ensure it is put to the service of creating public value. This situation demands developing and using a set of legal tools, organisational capabilities and digital public infrastructures that will enable cities to ensure that urban data benefits not only a few but society at large. This also requires capacity, policies, and skills to create and benefit from such public value.

In this report, we set out to combine insights from many years of data sharing experience with the specific insights gained from an experiment on urban data sharing in the City of Hamburg. On this basis, we have developed a range of recommendations. They are intended to enable cities and communities to access and use urban data to gain better democratic control of urban space and provide more effective public services.

1. **Shift the paradigm of urban data sharing: make access as broad as possible**

As always, change starts with a shift in mindset. The current paradigm around urban data still relies on the presumption that data should remain exclusively with the parties that control the technical access to the data, especially if these are companies. Data sharing happens under very narrowly predefined circumstances. Only a few new laws cautiously break this paradigm and introduce data sharing rights and obligations (such as the EU Data Act and the German Transportation Act).

To unlock the public value of data, a stronger push is necessary: Cities and decision-makers at all levels need to focus on the public value that can be created by sharing urban data. The presumption should be that urban data should be shared. In practice, cities should implement a policy that makes urban data available by default and use their legislative and contractual bargaining power to require private data holders to share their data in the public interest. They can do so in the following ways, for example:

1. in licence terms agreed with organisations that operate a service or run a business in the city,
2. in procurement agreements between the city and a contracting data holder,
3. in terms of public tenders, and/or
4. as a condition for obtaining public financing or public funding.

2. Give legal certainty: address protection laws

Legal uncertainty was a key challenge in the experiment in the City of Hamburg. Legal risks in urban data sharing stem from the existence of many and often conflicting laws that govern data sharing. Resolving these legal risks is currently often prohibitive for those involved in urban data sharing.

To systematically lower the risks, cities should consistently pursue an approach in which the interest in keeping data confidential is balanced with the public interest in making data accessible. Where possible, the need for confidentiality should be mitigated through technical–organisational measures, such as data aggregation. Cities can support this by working with the authorities that enforce protection laws to clarify which technical and organisational measures need to be taken to ensure legal compliance and who bears the responsibility and legal liability in each case.

3. Systematise solutions for urban data sharing: reduce sharing efforts

Sharing data requires major efforts by all parties involved because of the diversity of actors involved, the need for a common understanding of the urban data to be shared, and for a definition of the scenarios and rights of data use. These are currently defined from scratch in each sharing arrangement.

Cities should build and iterate a standardised use case repository by using a common framework that enables them to pursue a systematic approach to urban data sharing with a common language across use cases. This means identifying, developing, and reusing common ontologies, procedures, contracts, and templates that allow for systematising different sharing setups. For mandating or negotiating data sharing, cities should have and use standardised contracts and clauses with clear categories and terms.

4. Institutionalise urban data sharing: build a data intermediary

To make the complexities of data sharing more manageable, a specialised actor with an appropriate organisational and technical set-up can help address some of the challenges. Data intermediaries facilitate data sharing by defining and providing the most appropriate tools for it (legal, technical, or organisational). They can range from a data sharing contract between the city and other parties to a fully fledged organisation that handles data transformation and/or pooling, security, and other issues.

Cities should build urban data intermediaries to address the challenges involved. The most important tasks that such intermediaries can perform include taking efforts to assume any residual legal uncertainty and creating and sharing expertise on a joint standardised framework. They must follow a set of principles, including public purpose, not-for-profit, transparency, independence, inclusion, and accountability, to ensure that they serve the public-interest mission in urban data sharing.

5. Learn from Hamburg: Experimentation can be challenging but is essential

The New Institute and the City of Hamburg together defined and conducted a so-called challenge around urban data sharing. This process was considerably time and effort intensive for all stakeholders involved. The project brought together different parts of the city administration, the data protection authority, companies interested in sharing data, and organisations with ideas on how to use the data. It introduced innovation into the public sector at two levels: first, firms shared data with others for public interest purposes, and second, this happened in a challenge format previously unknown to the city.

Cities need to try out new ways of accessing, using, and governing urban data also when the final outcome is unclear. Introducing innovation inevitably implies that existing processes need to be adapted to make it work, which can be particularly challenging for city administrations. Resolving these challenges is vital for testing new ideas and generating new insights. Even if the processes used in an experiment are unsuitable for scaling, the learning experience will point in other directions to explore in the next iteration.

In summary, urban data can help cities to better serve the needs of citizens and become greener and fairer. Tapping into its potential must be a political priority, even if it entails legal, technical, and organisational changes and efforts. Cities must seize the opportunity with the measures available today to bring into practice urban data sharing in the public interest. We hope that this document paves the way for the discussion and implementation of these measures.

Table of contents

Foreword Francesca Bria

Director, The New Hanse

Europe can be a global reference for sustainable and democratic digitisation, combining the European Green Deal and Europe's Digital Decade and using the Next Generation EU funds and other large public investment programmes directed to the twin green and digital transition to make our cities greener and carbon neutral.

Building a smart city means designing the city of the future, which puts technologies at the service of people, starting from the great environmental and social challenges that cities are facing.

Cities in Europe can be the engine of change towards a green and ecological transition with social cohesion. Many cities are moving towards a vision of circular economy, investing in sustainable electric and hydrogen mobility, increasing the number of cycle paths and integrated public transport, with large urban projects such as in Hamburg, Paris, Barcelona, or Copenhagen, which free up public land and the spaces of the city centre from cars and traffic, with re-naturalisation projects to bring trees and greenery to the city, investing in renewable energy, making buildings more efficient and saving energy, making the waste recycling system efficient and circular. In this sense, there is an absolute need for cities to equip themselves with a digital infrastructure that collects public data on electricity and heat consumption, mobility, water management and pollution, developing new technological infrastructures, and data management as a common good, as a new urban public infrastructure.

It is a vision of the smart city that starts from citizens and the major urban, social, and environmental challenges to be tackled rather than from the technology itself: sustainable mobility, the fight against climate change, ecological transition, education, and health.

At the centre of this vision is the question of democratic control over the data. Data is the raw material of the digital economy, which is changing power relations in our society. Data in cities is a critical infrastructure, such as roads, air, energy, and water. It should be considered a public good because it is necessary to access collectively produced information while preserving citizen's right to privacy to make better decisions. Today, data is largely in the hands of private companies. That is why we launched The New Hanse initiative, a collaboration between the City of Hamburg and The New Institute to promote data sharing between businesses and society, transforming data into a common good that can be harnessed to make more informed public decisions and create public value.

The urban data sharing blueprint presented here, which was tested in the project, outlines the major legal, policy, and technological building blocks needed for safe,

ethical, and democratic urban data sharing that can be adapted, replicated, and shared among other European cities.

This report suggests smart and proactive policies, legal and technical tools, and new institutions, such as data intermediaries, that can shape the future of urban data sharing in cities across Europe and beyond. This blueprint shows how urban data sharing can work in practice. The prerequisite is mandating data sharing and access for the public interest.

It is about enabling citizens and communities to leverage data and digitisation to tackle our most pressing urban and environmental challenges, such as climate change, decarbonisation, and sustainable mobility. It is about forging a new citizen's pact on data. Now that Europe is passing important regulations, such as the Data Act and the Artificial Intelligence Act, hands-on projects, and experiments at city level, such as The New Hanse, can be game changers and pave the way forward.

Foreword Jan Pörksen

State Secretary and Head of the Senate Chancellery of the Free and Hanseatic City of Hamburg

© Senatskanzlei/Daniel Reinhardt

In today's digital age, data has emerged as a valuable asset throughout the economy and nearly all areas of life. For a modern city like Hamburg the smart use of (urban) data therefore also becomes increasingly important to render good public services to its citizens and enterprises. From developing new mobility solutions by optimising Intelligent Transport Systems (ITS) to formulating climate change strategies, the analysis and evaluation of data is a cornerstone for good policymaking.

In many policy fields, decision-making could be further strengthened if we were able to not just use the data we hold ourselves but also access privately produced and held data from companies and organisations. Accessing this valuable private urban data for the public interest in a reliable and constitutionally sound way is an intricate challenge though. We must delicately balance the legitimate interests of data holders on data protection, trade secrets and IT- security whilst demonstrating the envisioned purpose for the public good and fostering an innovative environment.

In the cooperative project between Hamburg and The New Institute, we have yielded valuable experiences on this model of data sharing for the public interest in practice and in theory. On the one hand, the concrete use case on micromobility flows through the St. Pauli district of Hamburg has provided us with a real-life testbed and shown the pitfalls and conflicting goals that arise on closer examination. On the other hand, through the two-fold set-up of the project, these practical knowledge gains could directly be streamlined into the theoretical work of the Data Commons Working Group and this blueprint.

Therefore, I personally look forward to the publication and the policy discussion on this blueprint and its actionable guidelines. As the Free and Hanseatic City of Hamburg, we look forward to reflecting its findings in the context of our own digital strategies and projects and encourage all other German and European public entities to use this blueprint as inspiration and action plan on how to access and govern data in the public interest and join us in our common way forward!

About The New Hanse

The New Hanse is a collaborative initiative between The New Institute and the Free and Hanseatic City of Hamburg, with a specific focus on business-to-government-to society (B2G2S) urban data sharing. The project aims to delve into innovative data governance models for digital cities. Central to its mission is the recognition that data stand as a fundamental utility for modern cities, a common good playing a crucial role in addressing urban challenges related to digitisation, climate change, and social cohesion.

However, a notable current limitation exists—much of the data generated by private companies in the public space, facilitated by sensors, phones, and other smart devices, remain largely inaccessible to city administrations, citizens, and the broader innovation ecosystem. Consequently, there arises a necessity for cities to explore novel approaches to govern urban data and strike a digital social pact. This exploration also involves the development of next-generation public digital infrastructures capable of seamlessly integrating data from diverse sources and platforms.

The New Hanse emerges within this context as a pioneering effort, seeking to explore the next generation of data sharing and data governance models for the public interest. It aims to enable B2G2S data sharing, marking a major step forward in how cities engage with and leverage data to benefit society. This work unfolds through two project components:

1. The Urban Data Challenge Hamburg:
This competition of ideas fostered experimentation in urban data sharing for the public interest. Focused on micromobility, it invited the innovation ecosystem to propose data-driven solutions, utilising shared private and public data to contribute to the city's green and digital transformation. The insights gained from this use case serve as an example and as a starting point for discussions within the second workstream, the Data Commons Working Group.

2. The Europe-wide Data Commons Working Group (DCWG):
Comprising experts from various fields and organisations (Annex 5: Members of the Data Commons Working Group), this group delved into data sharing regulation and governance within the Hamburg use case and beyond. Structured into legal, regulatory/governance, and technical streams, it developed actionable guidelines for European cities through interdisciplinary dialogue and collaboration.

Through the work in these two project components, **The New Hanse has accumulated valuable practical and theoretical knowledge on accessing, sharing, and governing urban data for the public interest** over the last two years. This report assembles the key lessons learned and condenses them into a policy toolkit with key recommendations for policy and decision-makers in cities and beyond.

About the authors

Francesca Bria (Director The New Hanse):
Francesca Bria is an innovation economist and an expert in digital policy, data and AI for the public interest. She acted as Director of The New Hanse at Hamburg-based The New Institute. She is a Board Member of the Italian public broadcast company RAI, Honorary Professor at the Institute for Innovation and Public Purpose at UCL, London, and is part of the high-level roundtable for the New European Bauhaus set up by the EC President Ursula von der Leyen. She is part of the Jury and a Senior Adviser on the EC programme STARTS (Innovation at the nexus of Science, Technology and the Arts). Since the year 2020 she served as President of the Italian National Innovation Fund, and she is the former Chief Digital Technology and Innovation Officer for the City of Barcelona, Spain. She has also served as Senior Adviser to the United Nation (UN-Habitat) on digital cities and digital rights. She led the DECODE project on data sovereignty in Europe.

Aline Blankertz (overall blueprint consolidation):
Aline Blankertz is an applied economist with years of experience in analysing and developing policy recommendations for platforms and data economy. She is an advisor on open-source and data policy at Wikimedia Deutschland and involved in various other projects, including those related to health and emission data. She has worked as an economic consultant and at a digital policy think tank.

Fernando Fernández-Monge (co-author and point of contact, governance section):
Fernando Fernández-Monge is a senior associate with the Bloomberg Harvard City Leadership Initiative, a joint programme of Harvard Business School and Harvard Kennedy School, funded by and executed in collaboration with Bloomberg Philanthropies. His work focuses on innovating within the public sector as well as fostering and managing technological innovation to make cities more efficient, equitable, and responsive to social needs. He holds an MPA from the Harvard Kennedy School and a Master in Fiscal Policy from UNED and the Instituto de Estudios Fiscales. He has written several teaching cases and academic and policy papers on urban governance, public innovation, and city leadership and management, published or featured in Apolitical, Governing, World Economic Forum, El Pais, and Agenda Publica, among others.

Joshua Gelhaar (co-author, technical section):
Joshua Gelhaar is a scientist at the Fraunhofer Institute for Software and Systems Engineering ISST in Dortmund as well as a PhD student at the TU Dortmund University. He has been involved in prominent data spaces initiatives, such as the International Data Spaces Association, Gaia-X, Catena-X, Mobility Data Space, and the Data Spaces Support Centre. His research focus is on data sharing and incentive mechanisms, especially within industrial data ecosystems. He holds a MSc in Industrial Engineering from TU Dortmund University.

Max von Grafenstein (author and point of contact, legal section):
Prof. Dr. Max von Grafenstein, LLM, is a legal scholar, lawyer, and serial entrepreneur. He has been working as an attorney since 2010 with a focus on intellectual property, internet technology, and media law, with iRights Law as a partner since early 2017 and since 2018 as an associate partner due to his appointment as professor at the Einstein Center Digital Future, Berlin University of the Arts. He also co-heads the research programme Governance of Data-Driven Innovation as well as the academic spin-off Law & Innovation at the Alexander von Humboldt Institute for Internet and Society (HIIG). His focus lies on the regulation of innovation with a particular view on data protection by design and data governance.

Ariane Haase (project lead, The New Hanse; co-author and point of contact, Urban Data Challenge section):
Hamburg-born Ariane Haase studied socio-economics and business administration and has more than seven years of international experience in consulting, prior to joining THE NEW INSTITUTE in August 2022. Ariane leads The New Hanse, holding all project strings together, and ensures alignment and knowledge transfer between the Urban Data Challenge and the Data Commons Working Group, blueprint drafting, and various project stakeholders.

Rainer Kattel (co-author, governance section):
Rainer Kattel is a professor and deputy director at UCL Institute for Innovation and Public Purpose. He has published extensively on innovation policy, its governance, and specific management issues. His research interests also include public sector innovation, digital transformation in the public sector, and financialisaton. His recent books include *Handbook of Alternative Theories of Economic Development* (edited with Erik Reinert and Jayati Gosh; Elgar, 2016) and *How to Make an Entrepreneurial State Why Innovation Needs Bureaucracy* (with Wolfgang Drechsler and Erkki Karo; Yale, 2022). In 2013, he received Estonia's National Science Award for his work on innovation policy.

Boris Otto (co-author, technical section):
Since 2017, Prof. Boris Otto has served as the director of the Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund. Since 2013, he has held the Chair for Industrial Information Management at the TU Dortmund. He is also a member of the boards of directors of the Gaia-X, European Association for Data and Cloud, AISBL, and the International Data Spaces Association (IDSA), and chairman of the board of directors of the Fraunhofer ICT Group. After studying industrial engineering in Hamburg, he did his doctorate at the University of Stuttgart and habilitated at the University of St. Gallen's Institute of Information Management, where he founded and managed the Competence Center Corporate Data Quality. His career path also included PricewaterhouseCoopers, SAP, and the Fraunhofer Institute for Industrial Engineering IAO. Furthermore, he was a research fellow at the Center for Digital Strategies, Tuck School of Business at Dartmouth College, New Hampshire, USA. His research focuses on industrial information management, business and logistics networks, and methods for the design of digital business solutions.

Oleguer Sagarra Pascual (co-author and point of contact, technical section):
Oleguer Sagarra Pascual is co-founder and co-CEO of Dribia, a data science innovation studio based in Barcelona. Dribia specialises in data strategy and design, implementation and maintenance of tailored algorithmic solutions that use advanced analytics and machine learning to understand, predict, and optimise business processes for the private and public sector. He holds a MSc in computational physics for UPC. His PhD focused on Network Science and human mobility at the University of Barcelona and included a stay at MIT working on research in urban shared mobility solutions. He keeps his research connection by participating in research projects, such as H2020 EU project DECODE, where he acted as pilot technical coordinator.

Lion Rackow (co-author, Urban Data Challenge section):
Since January 2022, Lion Rackow has been the Project Coordinator at The New Hanse project at THE NEW INSTITUTE. He was involved in the conception and implementation of the Urban Data Challenge Hamburg and the coordination of the Data Commons Working Group. Academically, he specialises in democratic theory and social change within the field of social sciences and holds an MSc in Political Science from Copenhagen University.

Section 1: Data Sharing for the Public Interest: urban data as commons

An introductory framework to the blueprint

By Francesca Bria and Aline Blankertz

Cities are laboratories for democratic and sustainable innovation: they enable experimentation at scale with new collaborative practices and democratic standards for data sharing as well as also wield normative and regulatory power coupled with infrastructural capacity and close connection to citizens.

The urban space—streets, sidewalks, parks, and parking lots—is where citizens meet and interact. With connected mobility and sensors for temperature, pollution, and noise, important parts of the urban public realm are becoming digital. This poses new questions: Who should control digital urban infrastructures, and how? How can data produced, collected, and stored in public space be controlled and governed democratically?

Without detailed information on what happens in our public spaces, we will not be able to tackle effectively the big societal challenges that we face in the 21st century, most notably global warming, and economic inequality. **We need transparency, access, and democratic control over the digital urban public realm.**

Responses to these challenges start with data[1]: Data is vital to understanding but also shaping how citizens use the shared public realm, such as streets, sidewalks, and parks. Currently, that data is not used for the public interest, but accumulated by individual players who do little to nothing with it, and when they do, their use is focused on exclusive, commercial use only. However, there is huge potential for society to derive information and wealth from this data and use it for the public interest to make urban mobility cleaner and more equitable, improve access to parks and public spaces, and, in general, generate outcomes that benefit urban communities.

1.1. Why we need to govern data as commons for the public interest

Cities need to leverage the potential of data for the public interest, i.e. for society at large[2]. This requires more data sharing, first and foremost with city administrations, who need data to better shape public spaces according to the needs of citizens. Cities have developed capacities to collect data, for example, on mobility flows and air quality, but this is not enough: many companies use the urban space and the data generated from that use exclusively for their own interests. This data needs to be made available to cities and others, including researchers, non-governmental organisations, and citizens who use the data for the public interest.

Urban data should be open and free to use for society unless there are good reasons to keep it private: it should be turned into a data commons to serve as a public infrastructure. Data is a key urban infrastructure, alongside roads, electricity, water, and clean air, and can be used to reach better, faster, and more democratic decisions, incubate innovation, improve public services, and empower people.
This means that the value of data should be accessible to society at large, while the state should ensure the data is accurate, of good quality and well maintained (directly

[1] Francesca Bria on data democracy: https://thenew.institute/en/media/data-democracy
[2] A new data deal: the case of Barcelona https://www.ucl.ac.uk/bartlett/public-purpose/publications/2022/feb/new-data-deal-case-barcelona

or indirectly). Data, similar to physical infrastructure, is a crucial input into many things that can be built on top. We need streets, sidewalks, and train lines to facilitate physical mobility. In the same vein, urban data can facilitate the detection of currently unmet citizen needs and the effective monitoring and steering of mobility to more environmentally friendly options. This data should be governed as a commons, i.e. by involving different groups including public authorities, local urban communities, and the private sector involved.

Urban data can be collected by public or private entities and can be shared not only with the government, but also with society more widely. These forms of broad data access shift the perception of data value: instead of focusing on the data itself, attention turns to the actual processes, results and products built from data because this is how value is generated. This broad data access accelerates innovation and levels the playing field. To ensure that societal goals are at the centre, conceiving urban data as a commons requires the participation of citizens and the involved stakeholders in defining and articulating what public value and interest mean[3].

1.1.1. How to read this blueprint

This blueprint provides guidance to municipalities on how to achieve the aim of treating urban data as a (data) commons serving the public interest. This section explains the challenges that stand in the way of such use and introduces the terminology we need to develop responses to these challenges. In the following sections, we explain:

a) the **legal** reasoning that underlies data commons and urban data access in the public interest (Section 2)
b) the **organisational and governance** set-up municipalities can introduce to enable an urban data commons for using data in the public interest (Section 3).
c) which **procedural and technical** steps municipalities should follow to build a robust infrastructure to access urban data for its use in the public interest (Section 4: How to implement and scale urban data sharing), and
d) what we can learn from the implementation of the Urban Data Challenge in the City of **Hamburg** in which businesses shared data on micromobility for public interest purposes (Section 5: The Urban Data Challenge in Hamburg: What works in practice).
e) Annex 1: Use case repository examples presents a framework for collecting use cases as well as example use cases.
f) Annex 2: Technical annexes presents further technical elements that can facilitate a systematic approach to data sharing.
g) Annex 3: Data sharing FAQs for cities: What elements are required when? lists questions and answers to guide cities on what is needed in different data sharing settings.
h) Annex 4: What we learned about using the challenge format for our experiment presents learnings from the challenge format used in Hamburg.
i) Annex 5: Members of the Data Commons Working Group lists the members of the Data Commons Working Group that contributed to the development of this blueprint.

[3] Francesca Bria, Our data is valuable. Here is how we can take that value back: https://www.theguardian.com/commentisfree/2018/apr/05/data-valuable-citizens-silicon-valley-barcelona

This introduction sets the scene for the following sections. We recommend reading it in full before turning to the individual sections, which can also be read in isolation.

In the remainder of this section, we first define the key terms: urban data as data commons and public interest. Next, we describe the key challenges municipalities face: an overly simplified presumption of confidential data, liability risks, and high transaction costs. We end the section by sketching out potential solutions to overcome the challenges. The following sections provide more detail on these solutions from a legal, governance, and technical standpoint. The identified proposals are based on an experimentation conducted with the City of Hamburg (the Urban Data Challenge) that will be documented in Section 5: The Urban Data Challenge in Hamburg: What works in practice.

1.2. How urban data should serve society

In this section, we clarify key terms including urban data, data commons and public interest as well as key actors involved in data sharing.

We refer to urban data as any data that is collected
(i) in the urban public space,
(ii) in the context of the provision of local services for the public via city procurement, or
(iii) in the context of city financing or public investment.

The public space is the infrastructure that is free for everyone to use, including streets, sidewalks, traffic lights, and parks. Local services for the public include street lighting, mobility services, garbage collection, and all services that, in principle, are accessible to all citizens. Both public entities and private companies collect urban data.

In this blueprint, we focus on urban **data collected by companies that have a public interest value** and the sharing of such data with the public sector and with society at large. For much of the data collected by public entities (authorities or state-owned companies), laws have been enacted in the last two decades to make this data openly available. This has created a more transparent and accountable public organisation. At the same time, it has also often exacerbated power imbalances with the private sector that has been amassing large amounts of data in the process of digitisation. Companies face no legal requirements to make this data accessible to stakeholders because no movement has sought to tap the considerable public value in that data yet. The value of such data lies in its statistical aggregation and contextualisation: We focus on statistically aggregated, non-personal data while recognising that the underlying operations may involve personal data. Data in scope of this blueprint is considered anonymised, i.e. the risk of personal identification is below the legal threshold.

Public debate around these issues is largely focused on the data itself (raw traces of processes), yet data only yields value when it is transformed into information and is therefore actionable. In the remainder of this blueprint, we distinguish between data (raw records) and information (aggregated or transformed data that is valuable as a basis for decision-making) and focus on the value generated from data.

1.2.1. What is the public interest

In this context, we understand the public interest as "those outcomes that best serve the long-run survival and well-being of a social collective (the public)."[4] Hence, public interest refers to outcomes that benefit a whole community and not just individual (personal or corporate) actors.

In the context of cities and urban data, focusing on the public interest means paying attention to **the collective value that can be generated from urban data** and the information derived from it. This focus also helps to avoid and change value-extractive practices where a handful of companies or individual profit (often exclusively) from data that has been generated using collectively held or publicly financed infrastructure. Therefore, to determine when there is a public interest, we need to identify the different ways in which urban data can generate collective or public value once it is made accessible. We also need to determine the best ways to govern data in the public interest – what are the needed regulations and institutions, and related public sector capacities, to ensure creation of public value through data?

1.2.2. How to leverage public value from urban data

Different information flows can be leveraged to create public value. Businesses can share data with the government (B2G), with other businesses (B2B or B2G2B) or with society at large (B2S or B2G2S). City governments in Europe[5] and the United States[6] have led pioneering efforts to mandate access to the data collected by micromobility companies, so they can enforce regulations and inform decisions around parking infrastructure or investments in public transit.

In some cases (B2G2B and B2G2S), data from different actors may need to be aggregated into information suitable to be used for its intended public. Such cases contemplate many-to-one-to-many information flow scenarios. In such cases, one (often public) entity gathers, processes, aggregates, and then shares the information to society as a whole or a subset of businesses in B2G2S and B2G2B cases. For example, in China, electric vehicle manufacturers have the obligation to share some of the data that they collect with a network of public centres processing the data, who in turn analyse and share back some of that data with other businesses as part of China's industrial strategy.[7] At the international level, an example of B2G2S can be found in the Net-Zero Data Utility, an "open, free, and centralised data repository that will allow all stakeholders to easily access key climate transition-related data, commitments, and progress of businesses and financial institutions toward those commitments".[8]

---

[4] Barry Bozeman (2007) *Public Values and Public Interest.*

[5] See the Dutch example available at: https://cds-m.com/, and the example of the city of Stockholm available here: https://miljobarometern.stockholm.se/content/docs/tema/trafik/elsparkcykel/Data%20driven%20regulation%20of%20micromoblity.pdf

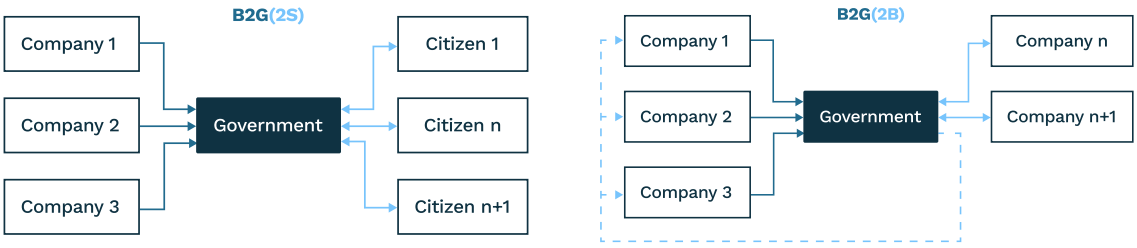
[6] See the US example available at: https://www.openmobilityfoundation.org

[7] See Martens and Zhao (2022) Data access and regime competition: A case study of car data sharing in China, available at: https://journals.sagepub.com/doi/full/10.1177/20539517211046374

[8] See https://www.nzdpu.com/

Although some of these use cases have been tested at both national and international levels, the complexity and challenges involved in such data sharing schemes have made them scarce, particularly at the local level. The New Hanse project was born precisely to experiment and evaluate how to make progress on this front at the urban level, linking it with the EU level.

Figure 1.1: Overview of data sharing scenarios



Urban data shared with the government (B2G)

The disclosure of information derived from urban data enables governments to generate public value by

(i) Better enforcing regulations in cities. For example, by accessing urban data on mobility assets (e.g. scooters and bikes) that are parked on the sidewalks, creating mobility obstacles, the city can enforce parking regulations.
(ii) Informing policies, services, and other decisions. For example, by accessing urban data on vehicle trips on public roads, the city can better understand where inhabitants travel to, how often they use mobility assets, or the average length of trips to make mobility decisions such as finding areas in need of further coverage or improving access to the public transport network.

For example, data collected by micromobility companies is increasingly required to be shared with city governments.[9] These mandates often include the establishment of standards and requirements (such as frequency and quality) for the data that needs to be shared between companies and city governments. Two examples of such efforts, including the applications that city governments can implement with the data collected from the private entities, can be found in the two initiatives, one led by a collaboration of cities in the Netherlands[10], and another one by the Open Mobility Foundation[11] in the US (and currently expanding to other countries).

Urban Data shared among private entities (B2B or B2G2B, when mediated by a public entity)

The disclosure of information derived from urban data among private entities can generate public value by:

[9] See Stephen Larrick (2022) Towards a Data Commons? On The Origins And Significance Of Platform Data Sharing Mandates, available at https://www.belfercenter.org/publication/towards-urban-data-commons-origins-and-significance-platform-data-sharing-mandates
[10] See https://cds-m.com/
[11] See https://www.openmobilityfoundation.org/

(i) Preventing market concentration: For example, data access to vehicle 'on-board diagnostics' by car manufacturers can ensure that more companies can provide services to car owners.
(ii) Promoting innovation: For example, the disclosure of data regarding the performance of batteries from electric vehicles with other manufacturers can help the industry in identifying the best solutions and foster rapid innovation in the electric vehicle sector.
(iii) Enhancing market efficiency by presenting a clear picture of its state while preserving individual business interest: For example, disclosing information about trip schedules or availability of ways of transport (e.g. in Mobility as a Service [MaaS] platforms; Annex 1: Use case repository examples), or publishing aggregated market data (e.g. total increase in scooter usage or total market share of small businesses) without identifying individual business performance (Annex 1: Use case repository examples). These are examples of a B2G2B case, since a central (often public) actor may have access to privileged information prior to disclosure.

An example of this effort has been launched in China, where the government established data sharing mandates to electric vehicle manufacturers. Such data is in turn collected, aggregated, and analysed by a network of public data processing institutions, which then make some of the processed data available again to various private actors to foster innovation. In this case, the data sharing mechanism is one of the tools of the broader industrial and innovation policy deployed by the Chinese government.[12]

Urban Data shared with society (B2S or B2G2S, when mediated by a public entity)
The disclosure of urban data with the broader society can generate public value by enabling the tackling of societal challenges. For example, the disclosure of data on pollution levels collected by private companies can be used by researchers to identify impacts of certain road characteristics (speed limits, lay out, natural barriers) on health outcomes in the city.

Urban data from different sources (e.g. navigation and location data from technology companies and other mobility operators, including bus and transport mechanisms) could be accessed and combined to develop mobility services for groups with limited accessibility such as older people, or to optimise people's and goods' mobility by integrating services to reduce CO_2 emissions and congestions.

An example of this in the urban context can be found in the NYC Recovery Data Partnership launched by the City of New York. The city requested data from different private entities to improve the understanding of COVID-19 impacts. On the one hand, this data helped inform programmatic and policy decisions, but some of the pooled datasets were also made available by the city government through its open-data portal. Such information was in turn used by different civic organisations to design "mobile friendly maps and community databases featuring the accessibility of food distribution locations, opening hours, and senior hours".[13] Another example of such a use case is

[12] For a detailed analysis of this case see Martens and Zhao (2022) Data access and regime competition: A case study of car data sharing in China, available at: https://journals.sagepub.com/doi/full/10.1177/20539517211046374
[13] For more information, see https://www.nyc.gov/office-of-the-mayor/news/542-20/mayor-de-blasio-launch-the-nyc-recovery-data-partnership-open-call-data

the Hamburg Data Challenge (Section 5: The Urban Data Challenge in Hamburg: What works in practice) which provided key lessons and insights for this blueprint.

1.2.3. Who is involved in data sharing

In the remainder of this blueprint, we refer to different roles that are needed for urban data sharing. Those roles are:

- **Data holders**: They have de facto power over data through technical means. They can usually exclude others from using data and can provide access to data. In the context of urban data, data holders can be public entities, however, they are often obliged to provide access e.g. through APIs or on open-data portals. Companies that are urban data holders often use public infrastructures to provide services in the public space and/or for the city.
- **Data contributors**: They are data holders that share data with other parties, for example, the government or an intermediary. Data contributors may decide to share data voluntarily or share data to comply with legal obligations.
- **Data users**: They benefit from the information generated from data for a particular use case, possibly by transforming and/or pooling the data. Data users can be administrations, researchers, companies, not-for-profit organisations, or citizens.
- **Data intermediaries**: They facilitate data sharing by defining and potentially providing the most appropriate tools for it (legal, technical, or organisational). Data intermediaries ensure data flow from data contributors to data users, for example, by clarifying the conditions for data sharing, transforming, or even pooling data. “Data intermediary" is a broader term than "data intermediation service" in the Data Governance Act. A data intermediary can range from a data sharing contract between the city and other parties that specifies how and what information can be accessed and for what purposes it can be used to a fully fledged organisation that handles the data transformation and/or pooling, information exchange, security, and other issues.

1.2.4. How to overcome the challenges

We identify two key challenges that need to be addressed to enable the use of urban data for the public interest:

- First, **protection laws:** The legal framework respects the right of individuals to protect information that relates to them as well as the right of companies to keep their business secrets confidential. There is high legal uncertainty on how to respect these rights when sharing data. The fragmented and diverse legal frameworks set up around data lead to individual actors struggling to understand and navigate those.
- Second, **sharing efforts:** Sharing data is associated with considerable and evident transaction costs. Compared with the more diffuse and hard-to-quantify benefits of data sharing, these transaction costs result in under-sharing, even when interests are aligned.

In the following, we sketch out how to address these challenges and enable the access, sharing, and governance of urban data commons.

1.2.5. Shifting the paradigm: Making urban data access as broad as possible

The general presumption that informs the current paradigm around urban data is that it should remain with the data holders as the party owning the technical means through which the data has been collected. This presumption currently applies even if other parties are involved in the generation of the data, for example by using mobility services (as citizens) or because they provide infrastructures such as public roads or subsidies (as city administrations).

As current practice shows, data holders usually use their technical control to exclude others and keep data confidential, despite the joint production of the data. This leads to a strong under-sharing of data which is at odds with the public interest for several reasons:

- First, companies often prefer to maintain exclusive access to data they collect because they fear losing a competitive advantage from it and because setting up data access is costly. However, data may also be shared without companies suffering a competitive disadvantage and without them having to bear the costs of sharing.
- Second, companies often block access to data even from those parties who have themselves helped to generate the data (e.g. by providing the public infrastructure or by using the service through which the data is generated). This is in conflict with the legitimate interest of these parties to also benefit from the value of the jointly generated data.
- Third, data access levels the playing field on which competition takes place because it constitutes a common base on which multiple organisations can develop innovation (such as new mobility services), thus avoiding winner-take-all marketplaces and the abuse of dominant position.
- Fourth, considering data an asset that should not be shared is an important blocker for using it within the companies themselves. This effectively hinders their innovation and potential to generate value. Having data ready to be shared externally also means having it in good shape to generate information for internal processes, for example, to scale processes and bring value.

New legal rules start to cautiously break with this paradigm: For example, the EU Data Act allows users of a connected device to request data generated by their use of the device to be shared with a third party (while, still restrictively, allowing data holders to hold back data by referring to business secrets and to avoid competition). The Data Act also gives governments the option to request data access under certain, fairly narrow, circumstances. Moreover, the German Transportation Act stipulates that certain (non-live static) mobility data needs to be accessible for a broad range of mobility service providers. In this case, the data owner is protected from competitive disadvantages insofar as the obligation to publish the data applies equally to all competitors. However, the Data Act will only enter into force from early to mid-2025, and the German Transportation Act does not have sanction mechanisms which means compliance is low. Furthermore, both laws only apply to particular types of data, while their conceptual basis of enabling other parties to use the data for their legitimate interests as well goes far beyond the types of data mentioned there. **This is especially true for**

urban data, which is not only generated by the use of public infrastructures, but in its use there is also a high public interest.

A stronger push is necessary to enable the use of urban data for the public interest: The value that can be created by sharing data should be put at the centre, while confidentiality can be mitigated through technical-organisational measures (e.g. data transformations, see Section 4.1.2). Only if the confidentiality interests may not be sufficiently safeguarded through technical-organisational means and these remaining interests still outweigh the public interest, this should stand against such sharing.

1.2.6. Give legal certainty: addressing protection laws

Legal risks stem from the existence of many and often conflicting laws that affect information sharing, including trade secret protection, competition laws, the General Data Protection Regulation, the Data Governance Act, the Data Act, transparency obligations, and sector-specific regulations. From the perspective of many data holders and data users, these legal risks are often prohibitive. The legal risks also apply to mandatory data sharing because, usually, the obligation to share data includes provisions that require compliance with such laws or at least not clarify their specific implementation.

To meet this challenge, it must be clear:

1. what technical and organisational measures can be taken to ensure compliance with protection laws, e.g. what data transformations are required to reduce the amount of information exchanged to a permissible level (see Section 4: How to implement and scale urban data sharing for more detail),
2. who must implement these measures, and
3. who bears the responsibility and legal liability in each case.

Hence, sharing and using urban data in the public interest requires complying with legal requirements. For this to be possible in practice, those legal requirements need to be mapped onto data transformations and aggregations to indicate which transformations are necessary to achieve legal compliance in each case (Section 4: How to implement and scale urban data sharing). This mapping requires a standardised understanding of legal compliance achievable by individual data transformations.

Additionally, there are several supporting processes that must be taken care of to generate the necessary trust via organisational and technical security measures. Those processes are the recording of all changes and steps of the process (logging), identifying who performs each change (identity) and the capacity to validate that all actions are according to contracts (audit).

1.2.7. Systematise and build on existing knowledge: reducing sharing efforts

Sharing data currently requires significant efforts, most of which are exerted from scratch for each transaction. We can distinguish three linked challenges that make these efforts necessary:

1. The diversity of actors to talk to and difficulty of reaching them, and the opportunity costs generated from their decision taking time caused by...
2. ...the lack of common understanding and identification of key functionalities, roles, actors, and checklists needed to gauge the effort, gains and consequences of the information sharing to take place which leads to...
3. ...enormous efforts spent on defining sharing terms, scenarios, and associated rights of data use.

Urban data sharing would hugely benefit from two solutions which, ideally, should be combined:

First, **a data intermediary** that, as highlighted above, facilitates data sharing between data contributors and data users. Cities should put data intermediaries in place, the governance of which we propose in Section 3: How an urban data intermediary should work.

Second, **a standardised framework** that identifies common ontologies, procedures, contracts, and templates that allow for systematising different sharing setups. Hence, cities should build a general unified framework and refine it iteratively by applying it to actual cases of data sharing. For the negotiation of data sharing, cities should have and use standardised contracts and clauses with non-ambiguous categories and terms, ideally in a machine-readable format akin to that proposed in other data sharing mechanisms (such as Gaia-X and IDSA). The continuous use of such a framework should give rise to a use case repository of reusable tools to implement across cities. We propose a complete framework that should serve as a starting point in Section 4: How to implement and scale urban data sharing and we further provide some basic examples of its implementation that should help in building a common repository of reusable cases in Annex 1: Use case repository examples.

1.3. What cities need to do to mandate data sharing for the public interest

Many cities are motivated to use data for the public interest. In the context of this project, we have worked closely with the City of Hamburg to experiment new forms of data sharing and data governance models on the ground. Although we recommend setting up a data intermediary for practical reasons (see below), cities can and should also take other steps to get more access to urban data especially from companies, putting in place the necessary regulatory measures and technical infrastructures to be able to manage this shared data as a common, and govern them in the public interest.

1.3.1. Which options exist for enabling data access in the public interest

Municipalities have several options to create mechanisms (mandatory and voluntary) for data access rights, enabling the sharing and governing of urban data commons in the public interest. We expand on them further in the legal section below. In short,

Municipalities can require private data holders to share their data:

1. when private operators seek to obtain a licence to operate a service or run a business in the city (e.g. shared scooters),

2. a data sharing clause may be established in procurement agreements between the city and a contracting data holder,
3. a data sharing obligation may be introduced under the terms of public tenders, and/or
4. a data sharing obligation may be made conditional to obtain public financing or public funding.

1.3.2. Why a data intermediary has great potential for the public interest

A data intermediary can implement or facilitate the implementation of the mechanisms and steps proposed above. It can provide the infrastructure for different parties to share their data in a standardised way, reducing the need for developing mechanisms each time from scratch. It can ensure remaining risks are well distributed between parties and provide a high degree of legal certainty. It can also enforce data sharing requirements for private actors who operate in the urban space. Finally, it can also provide access to society who can use the data for the public interest alongside the city itself.

Although such an intermediary yields enormous potential to fulfil the promise of urban data sharing for the public interest, placing an entity at the centre of such a process, thus inducing power asymmetries on the system, also entails significant risks. Hence, not all governance and organisational forms are compatible with the objectives of urban data for the public interest. Special care must be taken in shaping what such an actor can and cannot be and do, how it is financed and controlled and other governance aspects. We expand on these points in the governance Section 3: How an urban data intermediary should work.

The adoption of intermediaries also represents a unique opportunity to create a federation of intermediaries with standardised vocabulary and grammar for data sharing to reduce the currently prohibitive efforts for urban data sharing. The legal, governance and technical aspects of data sharing need to complement each other to ensure replicability across use cases and the reliable technical implementation of legal requirements.

Section 2: Why and how to mandate urban data sharing

The legal perspective

By Max von Grafenstein

This section focuses on how to make urban data access as broad as possible to facilitate its use in the public interest. We explain why urban data should be legally embedded as a commons. Data commons means, in principle, that everybody has an equal right to generate value from data (i.e. the information contained therein). This equal right of use only applies in principle because the data contains different values and risks depending on various factors: the perspective of the stakeholder concerned (e.g. the respective user of the data, the holder of a secret, the individual about whom information is collected), the context in which the data is collected and how it is processed by whom and for what purposes. Considering data as a commons requires finding the best possible balance between the interests that arise when putting the data to use.

2.1. The necessity of mandatory data sharing

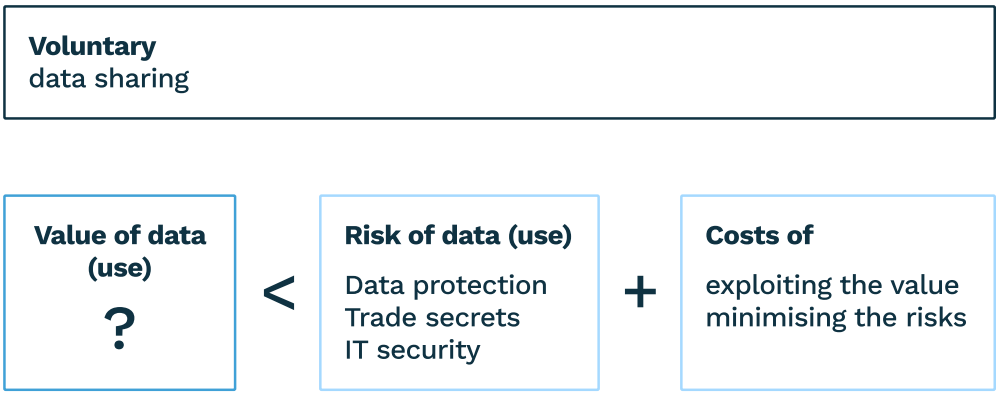
2.1.1. Why voluntary data sharing usually does not happen (at least not on a broad scale)

In principle, voluntary data sharing could be a good starting point: for voluntary data sharing to take place, data contributors need to assess for themselves what added value they see in sharing "their" data and whether this expected added value is higher than the perceived risks and costs. To convince a data contributor to share the data, data users may offer different incentives, such as:

- quid-pro-quo data sharing, i.e. access to other data in return (or even a pool of data, e.g. in the case of mobility-as-a-service platforms described in Annex 1: Use case repository examples);
- services for free (e.g. data transformations, analytical results, or a data-driven service);
- money or other financial benefits (e.g. beneficial regulatory or tax treatment for the data contributor, when possible).

However, as also demonstrated in the Hamburg Data Challenge (Section 5: The Urban Data Challenge in Hamburg: What works in practice), there are three decisive problems in voluntary data sharing cases: First, it is the data holders who decide whether data sharing takes place or not. This raises the normative question of whether purely technical control over access to the data should automatically give the data holder the legal right to decide on the value creation of the data alone (see in particular the following section). Second, the technical ability of the data holders to exclude others from accessing the data means that only, or at least primarily, their perspective determines whether sharing the data is worthwhile. This limits the value of the data to the one perspective of the data holder, even though the value of the data actually lies in its multiple usability by as many stakeholders as possible (who each bring their specific knowledge, skills, and opportunities to the value creation process). Third, data users are mostly not in a position to make a convincing value proposition to the data owner before they have received the data. This so-called value-for-risk dilemma alone leads to the fact that the data is hardly ever shared. This will be discussed briefly in more detail:

Figure 2.1: Voluntary approach to data sharing



The value-for-risk dilemma describes the situation in which a data holder only voluntarily shares "their" data with a data user if the user is able to offer the data holder added value that is greater than the risks and costs perceived by the holder in sharing the data. In some cases, the data user may be able to offer the data owner its own data or a certain service as sufficient compensation. However, in the vast majority of cases, a very abstract added value offered by the data user to the data holder is offset by very concrete compliance risks that the data holder faces when sharing the data. Especially in the case of data-driven innovation, the data user must usually obtain the data first to then explore a possible added value at some point in their innovation processes, in which they would have to let the data holder participate. Furthermore, even if the data contributor and data user are able to minimise the risks to a level where the expected value is higher than the risk, the effort required to do this (e.g. in the form of investments in infrastructure and technical capabilities) is usually so great that sharing the data does not seem worthwhile (see the Urban Data Challenge as a typical example for such efforts). In these cases, it is therefore also impossible to determine an objective market price that a data user may offer to the data holder due to its indeterminate value creation (not to mention the prohibitive risks and costs).

This is why voluntary data sharing does not achieve the goal of the data strategies of the EU Commission and the German government: To unleash the potential value of data on a broad scale.

2.1.2. How data sharing obligations or data access rights help

To exploit the value of data by as many stakeholders as possible by, at least partially, overcoming the value-risk dilemma, one way is to rely on mandatory data access rights for data users. In this case, the data holder's calculation no longer matters, according to which the value offered to the data holder by the data user must exceed the risks and costs in sharing "their" data. Thus, in mandatory sharing situations, the data holders must share the data without the data user having to let them participate in the added value of the data. At first glance, this approach sounds more drastic than it is usually implemented by the regulator.

The reason for this is that legislators have usually designed such data access rights or data sharing obligations with due consideration for the interests of the data holder and

affected third parties. For example, the current Data Act-draft of the EU Commission allows users of a connected device to request from the data holder to share data that has been generated through the use of the service with a third party. However, at the same time, the Data Act-draft widely respects the data protection rights of third parties as well as the data holder's interest in maintaining its business secrets (Art. 4 sect. 3, Art. 5 sect. 8, Art. 8 sect. 6, Art. 19 sect. 2) and avoiding a competitive disadvantage (Art. 4 sect. 4 and Art. 6 sect. 2 lit e).[14] This means that the de facto data holder may very well continue to generate value from the data for themselves, in particular by improving their own products or developing new products and thereby gaining or further expanding a competitive advantage. Mandatory access rights are therefore often designed in such a way that they largely respect the interests of the data holder. Thus, what such sharing rights primarily aim to achieve is that the data holder should no longer arbitrarily withhold the data and thereby prevent even non-competitors from using the data to generate added value, too.

Of course, such data access rights or data sharing obligations should not obscure the fact that the value-for-risk dilemma partly remains. The most important difference in mandatory data sharing situations compared with voluntary sharing situations is that the perspective changes and the balancing process becomes less complex: in mandatory sharing situations, it is no longer the data holder who weighs up and decides whether data sharing is worthwhile in terms of added value offered, as well as risks and costs, but the data user. Risks and costs also accrue for the data user because they often still have to respect data protection law and also the business interests of the data holder and third parties. In this situation, it is the data user who should ask themselves whether it is worthwhile asserting their data access rights which could even be contested by the data holder (e.g. on the basis of Art. 32 Data Act-draft). However, after all, the data user does not have to let the data holder participate in their added value, especially, not before assessing it in practice. This makes the data user's assessment of whether data access is worthwhile less complex than the data holder's assessment of whether they should share the data voluntarily. So this reduction in complexity is only one building block, albeit an important one, in unlocking the value of data. Of course, while sharing obligations make data access less complex, it is still risky: Without access to the data, a data user cannot know if the data can solve the problem at hand.

2.1.3. Why data intermediaries are needed also for mandatory data sharing

Another critical building block for fostering data sharing is the existence (i.e. establishment and support) of data intermediaries that help data contributors and data users to further tackle the complexity of data sharing. As shown, in all sharing situations, data holders and data users must usually comply with various overlapping (sometimes even conflicting) laws, first and foremost data protection, but also trade secrets and others (see again the Section 5: The Urban Data Challenge in Hamburg: What works in practice as a typical example for this). These laws must typically be

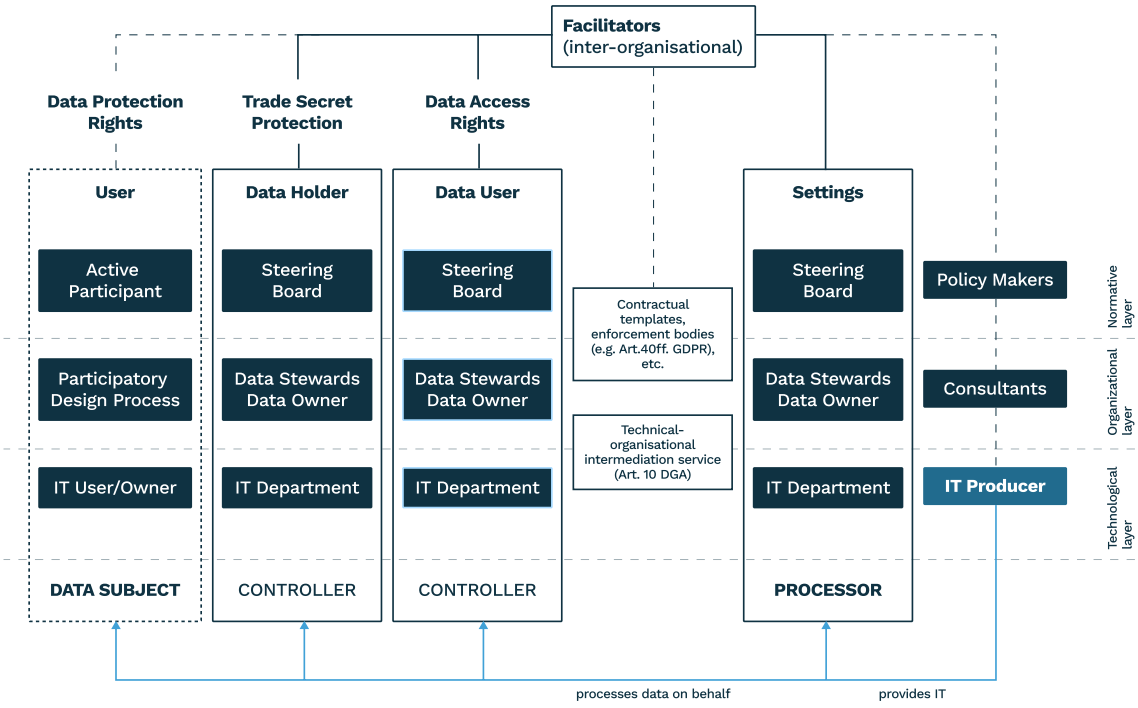
[14] The latter means that the data user must not use the data to develop a product that competes with the product from which the data originated, or pass the data on to another third party for this purpose.

implemented in the technical and organisational design of the data (sharing) processes. This task is complex for three reasons:

1. because the legal, organisational and technological governance layers are intertwined;
2. because of the dynamics of the value and risk assessments and the need to continuously adapt the measures on all layers accordingly; and
3. because of the resulting coordination effort for the people involved, given their different terminologies, goals, processes, methods, cultures, and so on (see the proposed common ontology in Section 4: How to implement and scale urban data sharing).

Data intermediaries can help address these challenges on all three data governance layers and take all forms, either as simple standard clauses or legal entities providing sharing services on all three layers, either in an integrated way or separately (which depends not only on economics of scale but also on the applicable law). The following chart may illustrate the functions of data intermediaries on all three data governance layers:[15]

Figure 2.2: Functions of data intermediaries



For instance, data intermediaries can help to coordinate the reconciliation of conflicting interests, especially at the organisational and technological level (see for example the function of "data sharing services" in Chapter III of the Data Governance Act). On the legal level, intermediaries may also contribute to increased legal certainty. This applies to the protection of trade secrets, where the settlement of disputes is in any case the responsibility of the data contributor and data user (through means of civil law). By

[15] v. Grafenstein, Reconciling Conflicting Interests in Data through Data Governance. An Analytical Framework. HIIG Discussion Paper Series, 2022 (2).

contrast, in data protection law, which is also enforced by public authorities and in some cases carries high threats of sanctions (see Art. 83 GDPR), data holders and users may also increase legal certainty through establishing and adhering to so-called codes of conduct and certification programmes (as foreseen under Art. 44 et seq. GDPR). Here, intermediaries may act as monitoring or certification bodies making sure that the sharing of data carried out on the technical-organisational layers meets the legal standards. Finally, regarding the Data Act-draft, it is worth discussing whether data intermediaries may or even should also function as dispute resolution bodies within the meaning of Art. 10,[16] given their expertise and a sufficiently guaranteed independent and neutral position. In conclusion, data intermediaries should actually do everything that helps data contributors and data users to share the data - nothing less (for more details on data intermediaries see the governance Section 3: How an urban data intermediary should work).

2.2. Implementing data commons through mandatory data sharing in the public interest

When implementing mandatory data sharing rights (for data users) and obligations (for data holders), public bodies (e.g. the legislator and municipalities) must respect the constitutional principle of proportionality. In a nutshell, this fundamental principle requires public bodies to weigh the interests of the data holder in keeping the data for themselves against the interests of data users and the public to use the data as well. This balancing act works like a seesaw: The more a data sharing obligation considers the interests of the data holder, the less weight the interests of the data user and the public may have to legitimise the sharing obligation. The other way around, the more important the interests of data users and the public in accessing the data, the more they may justify a stronger restriction of the interests of data holders. This ultimately means that the concept of the data commons does nothing more than implement the principle of proportionality by finding the best possible balance between the conflicts of interest in data. To this aim, the public bodies must counterbalance, with their legal means, the de facto power of the data holder by opening its technical access control over the data. The following comments highlight three aspects that are decisive for the outcome of the proportionality assessment for such legal means.

2.2.1. The interests of data holders and third parties in not disclosing certain information

First, it should be highlighted that the interference with the interests of the data holder[17] is fairly low, as far as a data sharing obligation or data access rights establishes one or more of the following conditions:

[16] According to Section 1 of this Article, "data holders and data recipients shall have access to dispute settlement bodies (...) to settle disputes in relation to the determination of fair, reasonable and non-discriminatory terms for and the transparent manner of making data available (...)."

[17] Which in most cases are at least protected by their fundamental right to conduct a business under Art. 16 ECFR.

1. the data holder is still allowed to use the data largely unrestricted on their own, i.e. for example to improve their own products or to develop new products and so on;
2. the data holder's business secrets are largely protected, in particular if the data user must not use the data to improve or manufacture products that compete with the data holder's products;
3. the data holder does not have to bear any costs incurred by sharing "their" data, i.e. in particular by implementing the technical and organisational measures that value creation and risk minimisation require.

The more of these conditions a data sharing law imposes on the data access rights of a data user, the smaller the restriction on the interests of the data holder. In the event that a sharing law even imposes all three conditions, the only thing a data holder may no longer do is withhold data arbitrarily, based solely on its factual control. Because of this minimal restriction on the interests of the data holder, there is a low legal threshold to justify such sharing obligations aka. access rights. Thus, any opposing legitimate interest by a data user or the public[18] may already justify such mandatory data sharing.

For this reason, it is possible that the legislator might even create a universal data sharing law that gives anybody access to all types of data, given that the data holder does not have to bear any costs and their further interests are respected (esp. In keeping their business secrets and so on). Only if the legislator establishes data access rights that amount to a more substantial restriction of the data holder's interests, a special justification would be required for reasons of proportionality. Such further justifications may be the role of the data user as co-generating the data or the overriding interest in using the data for a specific purpose (see the following two sections).

2.2.2. Empowering cities to use urban data based on public infrastructure

On the basis of the previous section, it is becoming clear that only if a data access right does not require the data user to fully respect the interests of the data holder, there is a need for further legitimising this access right. Such further justification may especially arise from the fact that a data user also generates the data. In this case of co-generating data, "co-generators" have in general the same legitimate interest and right in using the data. The reason for this is that the de facto holder of the data (who controls the technical access to the data) could only generate the data together with the other party. If both (or more) parties have generated the data together, there is therefore no legitimate reason to give one of them a better right to use the data than the other. Only because one party among the co-generators actually controls technical access does not give that party a better right, i.e. to use the data alone and exclude the others from using the data as well.[19] Co-generators therefore have the same right to

[18] Which is in turn protected by the data user's fundamental rights and/or the public interest as a whole.
[19] Special rights of one party may arise on the basis of intellectual property rights. However, as long as such intellectual property rights are not applicable to the case at hand, the purely

generate value from the data for any purpose they like, just as the de facto holder of the data.

Co-generation of data is, for instance, the conceptual basis for the data access rights in Chapter II of the Data Act-draft. Here, the legislator grants users of data-driven services the right to access data that has been generated through their use of the data-driven services. Because the users of the services co-generate the data (i.e. by using the service), the EU legislator puts the user at the centre of the regulatory system of these access rights. Consequently, third parties may have access to the data only through the user of the service (Art. 4, 5, and 6). Because the users of the data-driven services have helped to generate the data, they also have a right to generate value from the data (according to Art. 4 sect. 1 and Art. 5 sect. 1, the data holder must therefore also provide the data free of charge).

With respect to the public sector, the most important case for such an equal data usage right is urban data, i.e. data that is generated through the use of public infrastructure by the data holder. This means that if a data holder uses public infrastructure (e.g. gets a licence for using public roads or uses public spaces) or receives public money (e.g. in the context of public funding or public procurement), the public has an equal data usage right. The reason for this is, as explained before, that the data holder may not have been able to generate the data without the contribution of the public sector. Notably, in this context that Chapter V of the Data Act-draft, which regulates access of public authorities to data held by private bodies, falls far behind of the possibilities as explained here: Contrary to the access rights for private bodies in Chapter II, the access rights for public bodies in Art. 24 f. do obviously not refer to the role of the public as a co-generator. By contrast, these provisions only refer to the exceptional need of the public sector for using the data, too (see the following section). This is, indeed, a design flaw on the part of the legislator, leading to an unnecessary restriction of public access to such data generated through the use of public infrastructures, such as urban data.

2.2.3. Empowering cities to use other than urban data if there is an overriding public interest (or urban data under even better conditions)

Even if the data user is not helping to generate the data, access to the data may be justified on the basis of an overriding interest of the data user. A clear example of such overriding interests of private people may be their vital interest, for example when life and death are at stake. In this case, a data holder should certainly not be able to withhold the data, even if the sharing of the data involves the disclosure of trade secrets. With respect to the public sector, this is the conceptual basis of Chapter V of the Data Act-draft, which refers to the concept of "emergency" or "exceptional need" to grant public authorities access to data held by private parties (Art. 14 ff.). In such cases of overriding interests, one does logically not take into account (at least not comprehensively) the opposing interests of the data holder. This thought is reflected, more or less, in the following provisions:

---

factual control of data access does not give this party a legitimate right to exclude the other party from using the data as well.

1. Data access is unproblematic if the public authorities need the data to prevent a public emergency (Art. 15 lit a); in this case, the public authorities do not have to reimburse any costs for data access (Art. 20 no. 1).
2. If the authorities need the data to prevent a public emergency or to assist the recovery from a public emergency (Art. 15 lit. b), they must at least bear the costs for providing the data, in particular for anonymising the data, plus a reasonable margin. Here, the legislator obviously balances the lesser need for the data (compared with the previous case of necessity) by requiring the public data user to bear the costs and even pay a reasonable margin. In our opinion, such an additional payment is not necessary for reasons of proportionality. With this cost regulation, the legislator therefore went beyond what was required for reasons of proportionality favouring the data holder.
3. Finally, the same cost-bearing rule applies to the so-called backup clause (Art. 15 lit. c) sect. 1): Here, the public authority may access the data if it "has been unable to obtain such data by alternative means, including by purchasing the data on the market at market rates or by relying on existing obligations to make data available, and the adoption of new legislative measures cannot ensure the timely availability of the data". Thus, only if the authority could not purchase the data or (!) obtain it in time on the basis of a specifically created legal basis, it may invoke Art. 15 Data Act-draft.

This so-called backup clause seems unnecessarily restrictive when the data has actually been generated by the use of public infrastructure, such as in the case of urban data. In this case, the public sector should not only receive equal access to the data, but also not have to pay an additional margin for the provision of the data (but only the provision costs themselves, see the argumentation in Section 2.2). The public that has helped to generate the data by providing public infrastructure can thus use the data for every purpose they like as well and has not to justify this any longer with an overriding interest (such as exceptional need).[20] National legislators should therefore take advantage of the room for manoeuvre left by Article 15 of the Data Act-draft (which the Data Act legislator virtually points out to them) and create less burdensome data access laws for public bodies on their own.

This is all the more the case if the public sector may not only invoke its legitimate interest as a co-generator of the data, but also claim more specific public interests in using the data. For example, if a city has not only generated urban data by making public roads available (through which a data holder has collected the data), but also has a special interest in using that data for the fulfilment of a public task, e.g. for more effective traffic management, the city may demand the data from the data holder under better conditions than if only one of these two grounds were present. Thus, all three preceding grounds work like a pyramid: with each argumentative ground, it gets easier for the public authority to get access to urban data held by private parties.

[20] Otherwise, it would amount to a three-fold payment by the public: first, the public pays for the infrastructure necessary for data generation; second, the public pays for the sharing of the data (according to the Data Act); and third, the public pays for the prevention of an emergency in the public space, which not least also serves the interests of the data holders (who are usually also affected by the emergency).

2.3. Options for establishing mandatory data sharing

Roughly summarised, there are three ways for the public sector to create data access rights for itself and other actors. These result from the respective legislative competence.

2.3.1. Why a General Data Sharing Law (on the EU level) would be the best option

The most effective regulation would be the creation of a universal data sharing law. There are two reasons for this. The obvious first reason is that such a regulation would be the most comprehensive one: Everybody gets access to all sorts of data, as long as all protection rights of the data holder and third parties are respected (especially business secrets and data protection) and the data holder has to bear no costs (Section 2.2.1).

The EU legislator is certainly competent for such a comprehensive law, as the EU legislator may rely on the completion of the EU internal market. At least, the EU legislator has also based the GDPR on this competence. At first glance, one might doubt whether the data collected in the context of a Kindergarten, for example, has internal market relevance. However, since it may not be ruled out that at some point there will be a "smart" application that should collect the data of all Kindergartens in Europe, the EU legislative competence must be assumed on the basis of this potential value. In fact, the entire EU data strategy is based on this assumption of the potential value of data, that increases if the data is collectively produced, aggregated, and integrated.

Apart from the comprehensiveness of such a universal data sharing law, there is another, slightly more complex argument. Counterintuitively, such a universal data sharing law avoids the market distortions that sector-, context-, or even data-specific sharing laws would imply. The reason for this is that a more specific sharing law limits the data holder's data sharing possibilities: If there are no sharing laws, the holder of a certain type of data may exchange its data for other types of data (or completely different services) of a data user. However, if a data user receives access to this data on the basis of a sharing law, she does not have to give the data holder their own data. That this exchange possibility is important for data holders is increasingly shown by cases in which data holders from the public sector (e.g. public transport providers) have to make their data available to companies with predominant market power (e.g. large search engines or map providers) due to open-data laws without also receiving valuable data from these predominant market players in return. This market distortion widens with each sector-, context-, or even data-specific data sharing law. By contrast, a universal data sharing law would create an equal level playing field, because all data holders would be equally obliged to grant access to their data.

2.3.2. Establishing sector-specific data sharing laws based on (national) annex competence

Unless the EU legislator creates a universal data sharing law, the national

legislator may, of course, also enact a law to this effect. In Germany, such a law raises several questions regarding the legislative competence. In Germany, the legislative competence generally lies with the Länder. The federal government has legislative competence only in individual areas explicitly listed in the Constitution. Although this includes economic law, this competence is traditionally interpreted more narrowly than the corresponding legislative competence at the EU level with its "effet utile". Therefore, a federal competence for the creation of data access rights or data sharing obligations might therefore more often be invoked in the sense of a so-called "annex competence" (Annex-Kompetenz) or as a competence "by virtue of the factual context" (“Kompetenz kraft Sachzusammenhangs”). In this case, the federal government may only create data access rights if

1. the data is required for the preparation or implementation of subject areas that lie within the competence of the federal government
2. or, similarly, without which the subject area could not be meaningfully implemented.

The German Federal Transportation Act may serve as an example (Section 1.2.5). Here, the responsibility for regulating road traffic lies with the federal government (Art. 74 sect. 1 no. 22) and the federal legislator has obviously regulated the access or publication of corresponding data on the same constitutional basis. This example is not to say that the federal legislator, in the course of implementing its data strategy, may not at some point decide to enact a universal data sharing law based on its competence to regulate commercial law. However, the broader the scope of such a sharing law becomes, the more complex the legislative process and therefore unlikely the success of such a process will be.

Insofar as it is foreseeable that such sharing laws will also affect the transfer of personal data, the legislator should also create the necessary legal basis according to data protection law. This applies at least if the data is passed on to public authorities, since the consent of the data subject and the general clause of legitimate interests are not applicable here. In these cases, the legislator should therefore make use of the possibility pursuant to Art. 6 sect. 1 lit. e) and specify in these legal bases: especially, for which task carried out in the public interest the data is necessary; the types of data; the data subjects concerned; the specific entities to, and the purposes for which the personal data may be disclosed; purpose limitation; and storage periods.

2.3.3. The regulative power of municipalities

However, municipalities also have important possibilities in practice to create data access rights. In general, municipalities have the following options to require private data holders to share their data:

1. Data sharing obligations required to obtain a licence to operate a service or run a business in the city. The terms of the licence that these companies need from the city to run their operations or conduct business (e.g. shared scooters or electric vehicles) can include specific obligations on the types of data, their frequency and format, that they need to share with the city. Over the past years, local governments are increasingly adopting such data sharing mandates to digital sharing economy platforms deployed in their jurisdictions. Two German

cities have developed a practical guide on how cities can access a range of data.[21]

2. Data sharing clauses established in procurement agreements between the city and contractors. Also known as "data sovereignty clauses" and pioneered by cities such as Barcelona, these are standard clauses included in procurement contracts between city governments and their contractors. In these clauses, the city government establishes, as part of the contractual terms, the obligation by the provider to share data regarding the performance of the contract (e.g. service levels and unit costs) in machine-readable formats and on a regular basis with the city government.

3. Data sharing obligations under the terms of public tenders. Similar to procurement contracts, the terms and conditions of public tenders for different government services and infrastructure projects can include an obligation whereby the awardees of the tender need to share the data pertaining the relationship with the city government following certain criteria (including standards, quality, frequency, and format).

4. Data sharing obligations as a conditionality to obtain public financing or public funding. Public investment banks and other financial institutions owned or operated by public institutions can incorporate, as one of the contractual clauses or conditions of the financing instrument, an obligation for the recipient of the financing to share certain data with the financing institution or another public entity.

In all these cases, the municipality may rely on the balancing exercise outlined above (Section 2.2): the more the municipality protects the interests of data holders in these data sharing mechanisms, the easier it will be to justify the sharing mechanism. By contrast, the higher the (public) interests of the data users – whether because they helped generate the data and/or because they have overriding (public) interests in the data – the more they may constrain the interests of the data holders. This means that, depending on how high the overriding (public) interest is, they may restrict the data owner's trade secrets and/or make the data holder bear the costs.

To give an example: A micromobility provider as well as a map provider collect certain data from users on how they use these services in the public road space. A municipality wants to have access to this data to use it for more effective traffic management. In this case, the municipality (besides the end users of these services) not only has a legitimate interest to use the data as well, because they have generated the data by providing the public roads (and using the services). The municipality may also claim its specific interest in using the data for the public task of traffic management. In this case, the municipality may create a corresponding data access right through one of the options mentioned above. In terms of proportionality, the public authority will only disclose the data to third parties to the extent that this is needed for the fulfilment of the task (and implement appropriate protection measures to not unnecessarily restrict

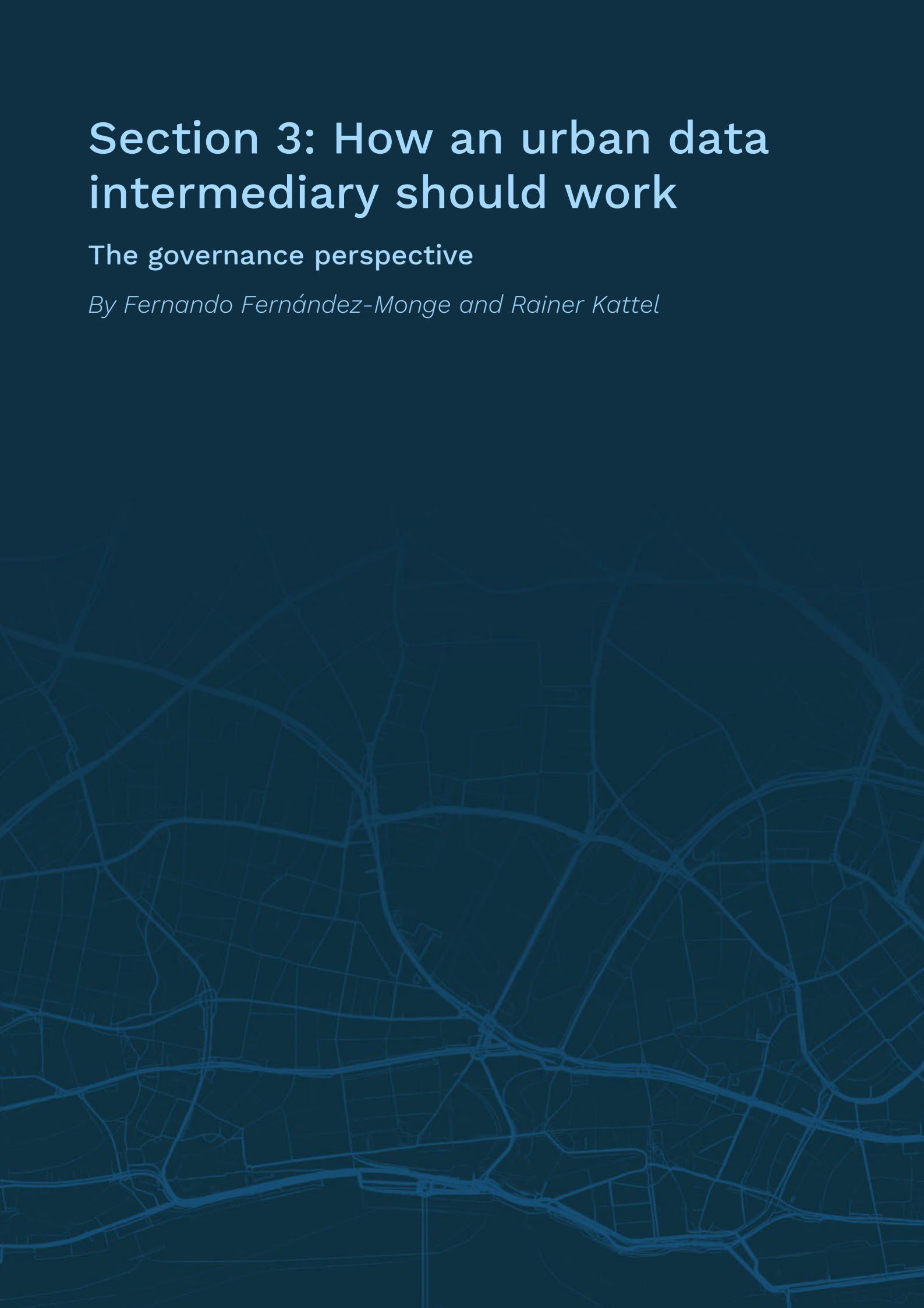
[21] See Praxisleitfaden für die Datensouveränität im Kontext von Open Data and the corresponding "source code"

the business secrets of the service providers and the protection of the personal data of the end users). To this aim, the municipality may also have to bear the costs for the necessary protection measures. However, due to its own exploitation interests, the administration will not have to bear the costs of the data holder for providing the data, and certainly not compensate the data holder financially (who will also benefit from the better traffic management implemented by the city).

Section 3: How an urban data intermediary should work

The governance perspective

By Fernando Fernández-Monge and Rainer Kattel



This section develops the idea of a data intermediary and its role in ensuring that urban data is used for the public interest. It describes the key areas that need to be considered when establishing such an organisation and presents some of the main options available. Although it has been drafted in the context of a collaboration project with the City of Hamburg, it shall also serve other cities and levels of government in Germany, across Europe, and beyond. Notably, each context, including Hamburg's, will have its own dynamics, necessitating more analysis and consultation prior to the final establishment of such an intermediary.

As mentioned in Section 1, "data intermediation" can range from a data sharing contract between the city and other parties that establishes the conditions of a data exchange to a full-fledged organisation that handles the data transformation and/or pooling, information exchange, security, value creation, and other issues. This section has been drafted considering the latter.

3.1. Why we need an intermediary

We have already seen that data sharing faces problems such as high liability risks due to a complex and constantly evolving legal landscape that is difficult to navigate or the high transaction costs that are derived from data sharing (discussed in Section 4: How to implement and scale urban data sharing). The data intermediary can help mitigate these challenges in the following ways:

First, the intermediary can help both data contributors and data users navigate the legal landscape by handling the responsibility of accessing, processing, and sharing the data. Given its concentrated knowledge and specialised expertise and skills, the intermediary can unburden parties, making it easier for them to exchange data and information while making sure that all the necessary legal requirements are met[22].

Second, the intermediary can help reduce the transaction costs involved in data exchanges. For example, the intermediary can collect the knowledge about different actors' and their needs and make this knowledge accessible to them, as well as establish a more direct and constant line of communication among them. Combined with the definition of sharing terms, scenarios, and associated usage rights described in Section 4: How to implement and scale urban data sharing, the intermediary's position at the centre of the exchange and its repeated involvement in data sharing transactions can reduce the frictions and time spent in data flows. A key tool in this regard is a use case repository developed and published by the intermediary. This will

[22] Delegating data stewardship responsibility to an organisation or group that is better equipped to make decisions already happens in other cases such as: the Clinical Study Data Request platform and the Yale University Open Data Access Project who steward clinical trial data on behalf of pharmaceutical companies; the Social Data Initiative that makes data held by Facebook available to social science researchers via an independent steering commission; and Genomics England's Access Review Committee which provides an independent evaluation of requests for access to genomics data. ODI (2019) Data trusts: lessons from three pilots. Available at: https://www.theodi.org/article/odi-data-trusts-report/

ease the discoverability, comparability, and analysis of urban data. Moreover, given its technical expertise, the intermediary can enable data exchanges in a secure, targeted, and controlled manner with full insight into who uses the data[23].

The intermediary cannot only tackle some of the challenges involved in data sharing but can also generate additional value that cannot be created in purely one-to-one data sharing scenarios. For example, the intermediary can create data pools that combine and aggregate individual datasets, and combine public, individual, and private data, enabling the generation of additional value (see, for example the examples of MaaS platforms or the data sharing for electric vehicle data implemented in China described in Annex 1: Use case repository examples). The intermediary can also provide additional services and applications leveraging the data it collects from different sources and providing useful information to third parties such as visualisation tools, information query services, predictive modelling, automated decisions, AI, and machine learning systems. The intermediary can provide assurance to the data contributors that the data is only accessed for the right purposes and can randomise the data that might be then published as open information. Doing so is a complex process so this can add a considerable amount of value both for data contributors as well as data users. Given its knowledge of the different actors, the intermediary can also do a more proactive and accurate matching of data holders and data users. This also provides additional incentives to certain data holders to join the data sharing scheme due to the participation of other data contributors who have complementary datasets. This, for example, is the case for MaaS platforms, where a neutral operator can overcome many of the challenges faced by current schemes (Annex 1: Use case repository examples).

Overall, a data intermediary can increase trust by citizens and stakeholders regarding the data flows and exchanges happening in the city. Given that much of the collected urban data has a social and collective dimension, this increase in trust is fundamental. As such, the intermediary should ensure that decisions about how data is shared and used occur following a social dialogue method, in a participatory, cooperative, and deliberative manner.

Nevertheless, data intermediaries are not a panacea and cannot solve all the challenges related to data sharing. Poorly designed data intermediaries can also create bigger challenges (indirect appropriation of data by corporates or dependencies by the public sector on data intermediaries) than the ones they are set to tackle. First, data intermediaries cannot be a substitute for investing in the required data governance and management capacities in public institutions. Creating intermediaries to merely address this lack of capacities and postpone or bypass strengthening is a short solution that will not solve long-term challenges. Ultimately, only a public entity with adequate capacities and responsiveness to its citizens will ensure that data is used for public value, and therefore, as described in the introduction (Section 1), the ability to create public value will also require data users with the capabilities and skills to generate such value. Second, the decision on whether an intermediary needs to be set up should be grounded in and informed by the conditions of each context.

[23] Sussa et al (2017)

3.2. When we do not need an intermediary

Because establishing an intermediary is not exempt from risks, its creation should not be an end in itself but rather a means towards achieving a goal. There may be contexts and circumstances where establishing an intermediary will not help achieve the goal of increasing urban data sharing for the public interest. Some of these circumstances are as follows[24]:

- When the potential harms of wider sharing of the data are very high.
- When the data that would be useful for achieving a public interest goal is not being collected or it is not collected in any useful way (e.g. not enough quality or timeliness of existing data).
- When data is already being shared effectively without the need for an intermediary (for example because the data is already or could be published as open data).
- When key stakeholders (mainly data contributors) are unwilling to share their data and cannot be asked to share the data in a mandatory manner; and
- When there is no political commitment or funding (this can be contested, because making funding available and allocating priority are political decisions).

3.3. Which principles an intermediary must stick to

After establishing the need for a data intermediary, the principles guiding the design and functioning of this organisation must be highlighted. These core principles are the following:

- **Public purpose:** The goal of the intermediary must be to enable the access of urban data from data contributors so that it is used for a public interest goal (as defined above). This is a principle that should guide the remaining features and rules governing the intermediary, such as who has access, to what data, for what purposes, on what terms, and for how long. It also affects other elements such as the composition of the governing bodies or the funding mechanisms, which should seek ensuring financial sustainability but not profitability (see next principle).
- **Not for profit:** The intermediary will be a not-for-profit organisation, and therefore earning a monetary benefit will not be a consideration informing decisions. The fees or other monetary exchanges that may happen while accessing and making data available for a public purpose will only be directed at ensuring the financial stability of the organisation.
- **Transparency:** Citizens need to be able to know what data is collected and made available by the intermediary (even when the data itself may not be publicly available). For example, the "logging" module described in Section 4: How to implement and scale urban data sharing mandates the storing of all interactions with the data (these, and not the data itself, might be shown to users). Moreover, the rules governing data access and use will be publicly available, as well as the decisions made by the intermediary during its operations. The same approach

[24] On the basis of ODI (2019) Data trusts: lessons from three pilots. Available at: https://www.theodi.org/article/odi-data-trusts-report/

will then need to be extended to the algorithms and AI systems generated with that data, following the algorithmic transparency principle now included in the EU AI Act.

- **Independence:** The intermediary will need to make decisions and act independent of political, business, or private interests, aiming to maximise the exchange of data to serve public interest objectives. Achieving this independence, however, can be difficult and will depend on the right design and management of the governance, funding, and decision-making processes to allow the intermediary to do so without any undue meddling by spurious interests. A first approach to the design of these processes is described in the remainder of this section.
- **Inclusion:** Civil society will need to be involved in the governance and decision-making process of the intermediary. The governance section later lays out mechanisms that need to be explored, such as the establishment of an advisory or supervisory board and the organisation of public hearings and consultations as integral steps in the intermediary's procedures.
- **Accountability:** The use of data to develop AI systems is a current and relevant concern; therefore, the intermediary needs to ensure that risks are considered and addressed by data contributors and data users and that the algorithms are accountable and explainable. To this effect, data contributors and data users that will develop AI systems using urban data will be required to declare (i) that they will do so, (ii) for what purpose (i.e. the ontology "use case" as described in Section 4: How to implement and scale urban data sharing), (iii) how they have incorporated ethical and responsible AI principles (in accordance to the EU's AI Act) to mitigate risks of biased and unethical outcomes, (iv) the metrics and the rules governing the algorithmics should be transparent and accessible to all relevant stakeholders

3.4. What an intermediary does

The intermediary accesses urban data (see Section 1.1 for the definition). They will identify areas of further refinement, such as frequency (static versus real-time data) and format (must be machine readable). Before granting data access, however, it must be established which data would need to be excluded from access (such as trade secrets or individual data protected by privacy and data protection rights). In addition, rules will need to be established regarding, for example, whether the intermediary will check the data quality and how a data contributor can cease to participate in the data sharing scheme.

Once data contributors share data with the intermediary, the access rights by data users will need to be determined. We can distinguish between different levels of access, which are still broad categories in a continuous spectrum; therefore, they will need to be made more specific in each use case. For example, in the Urban Data Challenge in Hamburg (Section 5: The Urban Data Challenge in Hamburg: What works in practice), some of the information (highly aggregated) will be made publicly and freely available to every potential data user. These can be made available with an Open Government Licence, or similar, which allows users to freely access, use, and

redistribute the data for various purposes, including commercial use[25]. Ideally, all of this information can be made available through application programming interfaces (APIs). This level can be determined as "low risk" regarding personal and business trade identification, as described in the technical Section 4: How to implement and scale urban data sharing (and developed further in Annex 1: Use case repository examples).

In cases such as the aggregated market data use case (mentioned in the Introduction and described in detail in Annex 1: Use case repository examples), more granular and detailed information will be made available (possibly through APIs to also access real-time information). Such access may require a subscription or licensing agreement, allowing users to access and utilise the information for a clearly pre-established public interest goal (including those specific commercial purposes that are aligned with a public interest goal).

In yet another set of cases, such as the Mobility as a service (MaaS) B2G2B, some of the information and insights may be made in the form of advanced visualisation or analytics products and AI-generated services delivered by the intermediary in exchange for a fee. In this case, rather than access to datasets or more granular information, the data user will get access to the information that solves their specific needs. Again, the access to such information and services will need to be justified from a public interest standpoint. Importantly, the process should ensure that data users cannot concatenate questions in a way that augments identification probability (Section 4.1).

Instead of looking at use cases, we can also distinguish between technical services that an intermediary must provide and those that it might provide as part of a wider portfolio. As mentioned in Section 4: How to implement and scale urban data sharing, the intermediary can rely on a technical data handler to carry out these tasks.

The minimum of services that the intermediary must offer relate to the management of the access to data. The intermediary should decide who can access the data and under what terms. This includes gatekeeping—that is, identification and authentication of data contributors and users during data flows. It also includes the identification of unacceptable data uses and the enforcement of decisions to stop and prosecute those unallowed uses. The intermediary should also carry out data anonymisation to ensure that the data from data contributors, when shared with data users, does not violate business trade secret protection and privacy protection (as detailed in the legal (Section 2) and technical (Section 4: How to implement and scale urban data sharing) section and Annex 2: Technical annexes for a more detailed description of anonymisation techniques).[26]

The intermediary can also provide other services that increase data availability and value. For example, the intermediary may engage in tasks related to storage and hosting, enabling them to check data quality and security or to pool and combine data.

[25] See for example, Transport for London's licencing terms, available at: https://tfl.gov.uk/corporate/terms-and-conditions/transport-data-service

[26] The anonymisation of individual personal data should be done by each data contributor before sharing the data with the intermediary. If they do not know how to do it, then the data intermediary can help, and in this case anonymisation becomes an additional service provided by the intermediary to data contributors.

This latter service is particularly important for some of the B2G2B and B2B2S use cases described in the Use Case Repository of Annex 1: Use case repository examples.

A core aspect for the implementation of the services and operation described in this blueprint is its public nature, and an implication serving from it is the need to therefore follow the free and open-source principle, along the lines of "public money public code." (which also forms the basis for the legal mandates of data sharing in public procurement, public tenders, and public financing described in Section 2. This will be key for the set-up described in this blueprint, and in particular its technical dimension, to iterate and evolve with the participation of a wider set of actors such as SMEs and start-ups. This is particularly relevant for smaller cities, who have more limited resources and therefore more susceptible to vendor lock-ins if the implementation of these services is enclosed. Moreover, promoting an implementation through the free and open-source principle will ensure that the transparency and accountability principle described above is followed.

3.5. Which organisational structure an intermediary needs

The intermediary can have several organisational forms, and the ultimate model will require further analysis, deliberation, and discussion among local key stakeholders. Yet, building on an analysis developed for the Dateninstitut (Data Institute)[27], two particular options seem worth exploring in more detail in the German context:

- A legal entity under public law with legal capacity (RF. AoR). An example of this could be the German National Library. Such an organisation would need to be established by a legal act and have a public mission. It can act with relative autonomy and flexibility when it comes to hiring personnel as well as establishing different funding mechanisms such as fees for services.
- A limited liability company (GmbH). Such an organisation has even more flexibility when it comes to attracting personnel (it operates outside the public sector personnel structures) and can provide competitive remuneration to qualified specialists.

Table 3.1 below includes an overview of the key pros and cons of each of the options identified by the Dateninstitut.

[27] Other options such as public corporations, foundations, etc. could also be explored but we focused on these two following the conclusions of the report commissioned by the Data Institute which assessed organisational forms for that Institute according to five criteria: difficult in incorporating the entity, independence, involvement of civil society, personnel, capacity to act. See report available at: https://www.bmi.bund.de/DE/themen/it-und-digitalpolitik/it-des-bundes/dateninstitut/dateninstitut-node.html

Table 3.1: Overview of organisational forms for intermediary

	Advantages	Disadvantages
Legal entity under public law with legal capacity (RF. AoR)	- High autonomy, only limited by ministerial legal supervision. - Representatives from civil society can be appointed to the Board of directors and committees. - Greater flexibility than public institutions to design their personnel hiring structure (who can have status of civil servants or mixed). - It may leverage the forms of action available under public law (administrative acts, contracts under public law), but also demand fees for financing the provision of services.	- Its establishment requires a legislative procedure. - The set up will be subject to a fair amount of negotiation and coordination- - It can take approximately 24 months to establish.
Limited liability company (GmbH)	- High degree of professional independence, limited by participation by public institutions on its governance bodies. - Representatives of civil society can be part of the board of directors, or indirectly via supervisory board or an affiliated not-for-profit foundation. - Open to all diverse types of employment and remuneration models. - Substantial ability to act with regards to service provision and financing.	- Its establishment has fairly high implementation costs. - It requires drafting of Articles of association, notarisation, and registration in Commercial registry. - It can take approximately 6–12 months to establish. - Less authority than under a form more closely associated with the State.

This information, and the report it is based on, can provide guidance on the criteria that can be used to evaluate the appropriate options.

In any case, the pros and cons of these two options, or other potential alternatives, need to be explored further before deciding on the optimal organisational form for the intermediary. For example, as discussed in the first legal working document preceding this blueprint, a private law entity that is not under the control of the City of Hamburg, the Hamburg Transparency Act does not come into play, and therefore the intermediary has more flexibility to determine the terms of data and information access. Another key consideration is to ensure that the determined legal form is deemed independent

to provide sufficient insulation to the city government from liabilities deriving from a complex legal setting and insufficient technical capabilities.[28]

3.6. Which governance a data intermediary needs

A key element of the governance of the intermediary will be the composition of its governing board. The board of directors will need to represent the different interests, concerns, and perspectives involved in urban data sharing. For example, the board could include a data governance expert, city representative, a community representative, an academic representative, and a business industry representative.

A sound corporate governance of the board of directors will be ensured by applying key rules such as diverse representation of interests, term limits, phased terms to ensure balanced succession, and conflict of interest guidelines.

In addition to the board of directors, the intermediary may have an advisory or supervisory board composed of known experts and recognised members of the community in Hamburg, who can provide general advice on the strategic direction and key decisions of the intermediary.

Lastly, the intermediary needs to develop different mechanisms to engage and interact with the broader civil society in the city as lessons are learned and experience is accumulated through the operation of the intermediary. Public hearings and regular public consultations need to be an integral part of the intermediary's governance structure and procedures.[29]
Of course, these are mere guidelines about how participation can be incorporated into the governance of the intermediary. The main goal is for participation to be at the front and centre of the design of the governance structures of the intermediary. This requires a more elaborate definition of all these aspects, detailing the particular participants, power-sharing mechanisms, and rules of decision-making and operation. Delving into these details goes beyond the scope of this blueprint; such details and definitions must be carefully tailored to the context of the intermediary during the design stage. The principle that such design should follow and fulfil is clear; however, participation cannot be merely residual—it must be a core element of the governance structure.

---

[28] See the legal working document by v. Grafenstein preceding this blueprint (sec. 2.3), available at https://thenewhanse.eu/en/a-first-legal-assessment

[29] Ibid. (sec. 2.3.3. p.22): *Beside an operational management body, the following structures and procedures may include An expert committee with the CDO, ITD, and BVM, the responsible freedom of information and data protection authority, and other experts. Furthermore, the participation of the data contributors and recipients, which is crucial for the successful identification and resolution of the conflicts of interest, in particular the FHH, the E-Scooter providers, as well as the E-Scooter users and other affected citizens (especially data subjects). These participants may be involved in one or another way in defining the data access and usage rules, in particular on the questions of a) which data recipients b) may access which data c) in which quality (e.g. aggregation) d) from which data contributor e) for which purposes f) under which technical/organisational conditions. The specific methods and the degree of participation, has to be specified in the following steps of the project in accordance with the specific question to be negotiated.*

3.7. How an intermediary should fund its operations

The key goal of the funding and revenue model should be to ensure that the intermediary can meet its goals in a sustainable manner. That means financing the functioning and operations of the intermediary with a long-term perspective.

The funding and revenue model should therefore not only consider current operations but also plans for future development as innovations and challenges emerge, with a forward-looking and adaptive approach. It is also important to diversify the sources of funding to increase the resilience of the financial health of the intermediary, as well as to think about how the intermediary can add value to data users in exchange of financial resources to sustain its operations, while remaining neutral and respecting its founding principles.

With these considerations in mind, there are several (non-mutually exclusive) revenue mechanisms available. For example, the intermediary can be funded from the government or other public organisations. This funding can also come from a set of different sources, including a combination of public and private (such as philanthropic organisations) funding streams. Another option is to generate revenue from services. First, notably, enabling data sharing is already a valuable service, because it demands the pursuance of legal and technical processes that, absent an intermediary and described repeatedly throughout this document, prevent data sharing for the public interest. Thus, if the intermediary provides data processing, anonymisation, aggregation, or visualisation of the urban data it collects, it may charge for these added-value services. Some organisations (both data contributors and data users) may be willing to pay for access to a support centre, training, or consultancy services that the intermediary could provide. The intermediary can also charge for API access and ancillary services, such as tailored feeds[30].

The fee structure should be fair and transparent and can be designed to give waivers or apply reduced fees to non-profit, academic, or other entities. Notably, according to the Open Data Institute, charging fees, particularly through licences to use data, can add transaction costs or restrict the flows and combinations of data, reducing the value that can be generated. If the intermediary has a central and monopolistic position over urban data, these transaction costs can become substantial and limit, rather than enhance, the access and use of urban data for the public interest. It is important, therefore, to ensure that any fees are well-justified and determined with the purpose of covering the costs of the sustainable operation of the data intermediary.

3.8. How the intermediary should assess impact and document lessons learned

The ideas developed in this blueprint are new and untested, so it will be important to constantly monitor, evaluate, and learn from the experience gained from the

[30] See ODI (2019) Data trusts: lessons from three pilots. Available at: https://www.theodi.org/article/odi-data-trusts-report/

establishment and operation of the intermediary. This should not be an afterthought, by contrast, a monitoring and evaluation framework should be set up from the start. This means determining the outcomes (e.g. public interest goals achieved such as particular local policies informed and improved thanks to urban data) and outputs (e.g. number of datasets collected or data users attended) that will be measured, in which timeline, and with what data and information.

A key tool in this regard will be the use case repository proposed in detail in Section 4: How to implement and scale urban data sharing. Such collection of use cases should not only document the design and implementation of data sharing scenarios. As these use cases are implemented, it will be key to measure, register, and publish key indicators regarding their impact. This will enable other policymakers to understand the public value that each of the use cases generates.

3.9. How to expand data sharing: a federation of data intermediaries

Although this blueprint was drafted in the context of Hamburg's pioneering and ambitious goals to increase the access and use of urban data for the public interest, it also seeks to be useful to other cities and beyond. This may mean expanding the reach of the intermediary beyond one city's limits.

There are at least two approaches to achieve this expansion:

First, the amount of data managed by the intermediary can be extended geographically. It can also increase the amount and types of data that it is accessed and used, including the incorporation of additional data contributors and data users. Such a scaling approach may not always be optimal, for example if it makes it difficult for the intermediary to fulfil its purpose or to remain effective.

A second option in the German or European context is to replicate the intermediary model in other – city or regional – contexts, creating connected institutions with similar statutes, governing boards, and data sharing rules. Of course, although the basic structure may be used, specific features of other intermediaries will need to be adapted to each context. In this blueprint we recommend this second option, that could operate as a federated network of intermediaries (also mentioned in Section 4: How to implement and scale urban data sharing). An example of such network of private data repositories managed by public entities can be found in the network of data monitoring platforms on New Energy Vehicles established by the Chinese Ministry of Industry through its National Technical Specifications of Remote Services and Managing System of Electric Automobiles[31]. A key role for such federation or association of intermediaries is thus to develop a common set of principles, standards, use case repository, and evaluation framework (as stated in the previous section), among other tools that can allow the interoperability that is necessary for the geographical scale-up of data and information sharing across jurisdictions. This federation of data intermediaries could be for instance tested in the context of European local digital twins.

[31] For a detailed description of this case see the Paper by Bertin Martens and Bo Zhao (2021) available at: https://journals.sagepub.com/doi/10.1177/20539517211046374

To overcome sharing obstacles across cities, such a federation of data intermediaries can leverage the technical procedures described in this blueprint, particularly with regard to its alignment to a legal framework that enables data sharing and its enforcement. Otherwise, if the standardisation is limited to legal aspects (contracts, certification schemes, legislative tools), and does not take into consideration its technical implications, it will be challenging to enforce and therefore very difficult to adopt. A key role of the federation of data intermediaries will therefore be to promote the standardisation of data sharing legal, but also technical, instruments and procedures (as proposed in the next section), and the proposal outlined in this document can serve as the building blocks for such endeavour.

Section 4: How to implement and scale urban data sharing

The process and technical perspective

By Oleguer Sagarra Pascual, Boris Otto, and Joshua Gelhaar

This section presents tools to scale the process of data sharing, in particular when supported by a data intermediary. We assume that the data intermediary exists with a proper governance and follows the guiding principles outlined in Section 3: How an urban data intermediary should work. We first identify three main subproblems of the sharing efforts that hinder the scaling of data sharing. We then propose tools to overcome them. Thereafter, we describe the technological architecture to support such tools. Annex 2: Technical annexes and Annex 3: Data sharing FAQs for cities: What elements are required when? provide expanded technical details and a practical checklist for cities, respectively.

4.1. How to reduce the effort required for data sharing

Many efforts have recently been made in the technological front to enable data sharing and use at scale, most notably around the International Data Spaces and Gaia-X initiatives[32]. These cases deal with scenarios where a single entity shares data with another (unknown) entity (one-to-one). By contrast, this blueprint deals with different situations where many known entities (mostly 1–10 private companies) share data with a single recipient (government), who may in turn share it with different data users (see Section 1) for an overview of different potential such cases).

Typically, the reasons why processes are not scaled[33] are not technological but lie in the difficulty to properly define them, so they can (later) be supported by technology, evaluated and enhanced[34]. B2G data sharing is no exception.
Current efforts across public administrations are not aligned, but use different definitions, contracts, criteria, and templates, effectively hindering the large-scale adoption of data sharing and the spread of good practices[35]. Three steps are helpful to align these efforts.

- **Description**: First, define properly the processes involved in data sharing scenarios, using unambiguous vocabulary and ideally quantitative measures (such as numbers or categories). The literature uses similar wording for different situations and objectives, causing much confusion. Without clearly defined tasks, actors, and decisions, each data sharing scenario looks unique, rather than a concrete implementation of a greater, underlying framework. In theory, infinite data sharing agreements can exist (including combinations of multiple actors, multiple access, and usage rights as well as multiple use cases). In practice, however, these agreements comprise a small, finite number of basic elements that are combined in recurring ways which can be identified and standardised to reduce complexity.
- **Mapping**: Second, ensure that this vocabulary can be used to describe effectively real cases. This means that if two cases are described equally, their analysis must

[32] See Otto, Boris, et al. (2021) GAIA-X and IDS. Position Paper.
[33] By scaling we mean, first, the vertical process of minimising human intervention necessary for low-added-value tasks, i.e. automating as much as possible, and, second, the horizontal process of maximising reusability, the continuous enhancement by repetition and establishment of patterns and good practices.
[34] See Accenture (2021) The answer is scale...what's the question?
[35] See Capgemini (2023). Connecting the dots: Data sharing in the public sector.

be the same, and critical decisions over them (such as legal rulings) should be consistent. As an example, we can find multiple cases where data protection authorities issue the same (or even different) rulings repeatedly over similar cases, without these rulings being publicly accessible to other parties. In the absence of knowledge about these rulings, the other parties must therefore start from scratch each time searching for a suitable solution.

- **Iteration**: Implement tools that support the identified data sharing processes so they can be reused and give rise to interoperable standards[36]. This means reusing data sharing terms and contracts, data transformations, procedures, and organisational tools across administrations. Work must also be done to ensure that people can find these tools and can adjust them to their needs.

We consider that these three steps, which come before technological adoption, are the key to widespread adoption of B2G data sharing practices across cities. In the following section, we make a specific proposal to operationalise each of these steps.

4.1.1. We need a unified technical vocabulary

To properly define the process involved in data sharing agreements, we define an unambiguous, clear, and shared vocabulary that allows us to untangle a seemingly broad concept as data sharing into smaller, well-defined processes with clear actors, tasks, and dependencies. This is what we call an ontology of the data sharing processes. Annex 1: Use case repository examples illustrates this tool with specific mappings of three use cases introduced in Section 1, one of which is the Urban Data Challenge Hamburg, a B2G2S case (Section 5: The Urban Data Challenge in Hamburg: What works in practice).

The ontology helps clarify the conditions of and speed up negotiations among parties to set up data sharing scenarios. When proposing a new data sharing agreement, each participant should use the ontology to understand what tasks it is expected to fulfil and what is needed from the other parties. This process can highlight any missing parts for effective data sharing and help in their planning.

The proposed ontology is not a specific proposal for implementation, but a general mapping tool. To define it, we start with three key principles:

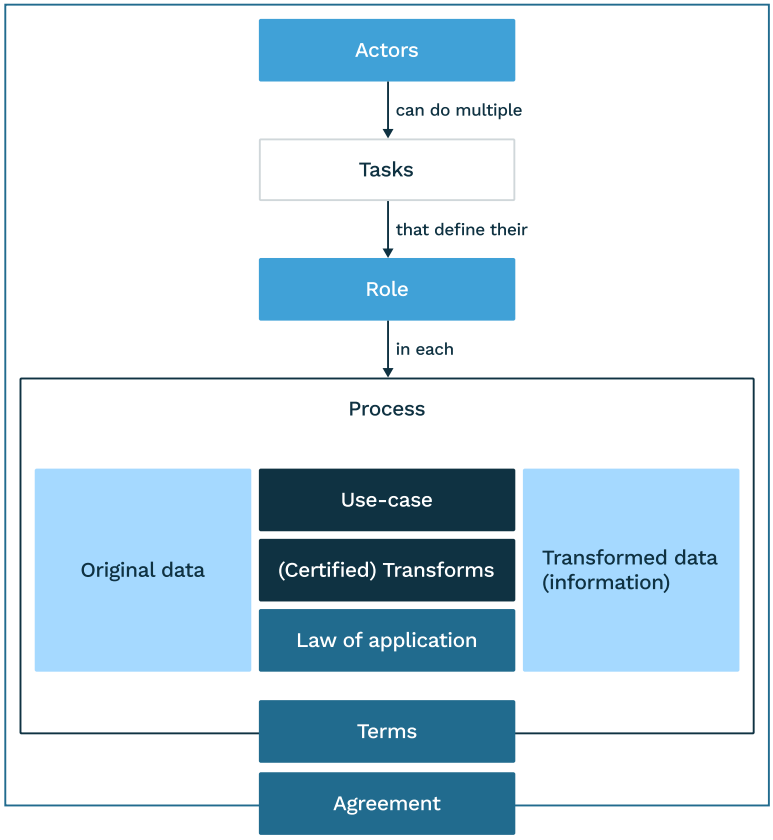
- First, data sharing scenarios are about the information generated, not the data itself[37]. A data sharing scenario, in a nutshell, is a set of processes by which data contributors provide data to a single data intermediary. The intermediary transforms data into information that is useful for data users for particular use cases defined beforehand. The use case determines which information is useful so information disclosure should be limited to fulfil this objective and data transformations adapted to that end[38]. This definition implies that different data users may receive different transformed versions of the same data, that respond

[36] Adoption of interoperable standards across cities and departments is key to success. That is why our proposal of data intermediary contemplates the establishment of a network of entities across cities that can share, update, govern, and enhance said standards.
[37] See Open Data Institute (2023) Understanding the social and economic value of sharing data.
[38] Thus, applying the principle of proportionality discussed in Section 2.

to their interests. This clearly implies a lack of a one-to-one mapping between data and information; moreover, each piece of information serves different uses and hence carries different risks that may legally require different data transformations.

- Second, one must be able to split any data sharing scenario into small processes composed of tasks that are carried out by a single actor. A process must contain two key roles (data contributor and data user) and can contain auxiliary roles.
- Lastly, all actors must be known and identifiable, and all actions traceable and logged, to fulfil the principles of transparency and accountability (see governance Section 3: How an urban data intermediary should work)[39].

Figure 4.1: Basic elements of the data sharing ontology



We sketch the basic structure of the ontology in Figure 4.1. Actors can be the city administration, businesses generating urban data, and potential users of the generated information such as (the same or) other businesses or other societal actors (NGOs, citizens, researchers), the entity acting as data intermediary, the data protection authority, and others. The data intermediary must help the different actors enter into broad governance agreements. Those agreements are composed of *sharing terms* (or terms) that spell out which actor takes which role (and for how long) in each process, considering what data is provided originally, and to which end (use case) it is required.

[39] The case where the ultimate product of the data sharing is meant to be open data may limit the applicability of this principle (data user may want to remain anonymous). In this case, we define "identifiability" by the fact that one is able to discern individual uses among the total, even if those are all anonymous (but discernible).

The use case determines which information is required and, hence, what protective laws must be respected via suitable data transformations. The agreements also need to define how those terms may be modified and terminated.

Data sharing scenarios can contain multiple data contributors and the data they provide should have as little personal identification risk as possible. The scenarios can also contain multiple data users.

In addition, auxiliary roles are optional and are not required but can help to reduce the risks and efforts in data sharing. These are:

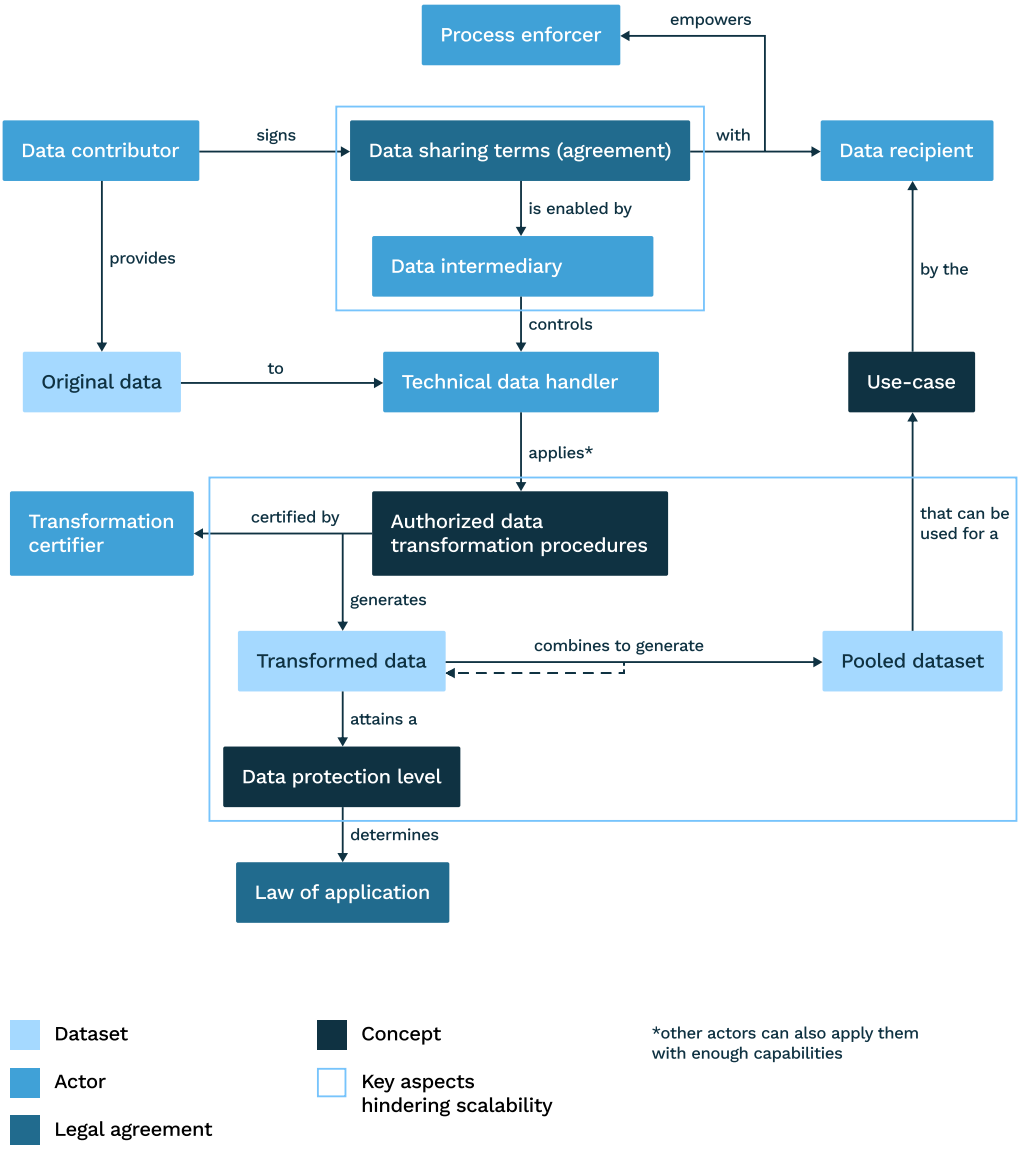
- The way the data is transformed from the original to the pooled dataset is determined using the agreed data transformation procedures among the actors, that can be certified by a transformation certifier agreed between the parties. We denote as a transformation certifier the entity in charge of determining what laws govern which datasets, and of validating the proposed data transformation procedures, aimed at satisfying a certain degree of leakage risk, both at the personal and business trade protection level. It can be a public authority such as a data protection agency.
- If the data intermediary itself, or any of the involved parties, do not possess the needed technological means to deliver the agreed datasets or implement the data transformations, they may externalise the treatment to a technical data handler.[40]
- To ensure the accountability and transparency of the entire process, all the parties can agree to nominate a process enforcer, an actor with the overseeing power to audit that all the tasks happening with the data sharing process are implemented according to the agreed atomic data sharing agreement.

In the above set-up, the data intermediary is a facilitator between data users and contributors to establish the terms of the agreement. The complexity of the data intermediary will depend on the needs of each case. It can range from a straightforward set-up (such as a written agreement) to a fully fledged organisation governing such an agreement and taking legal responsibility for the outputs (Section 3: How an urban data intermediary should work).

[40] Note that if the data contributors do not have enough technical capabilities, the task of preparing the original data can be made by the technical data handler. Certified transforms can be applied by both data contributors or data users if the parties agree to delegate this process.

Figure 4.2: Ontological elements of a data sharing process

The New Hanse_Blueprint_Governing Urban Data for the Public Interest



A full data sharing scenario will typically comprise different processes and, thus, different terms. For instance, a company can act first by providing their original data (thus acting as data contributor) that is later combined with other datasets into a pooled dataset and delivered to a third party (data user), that generates outputs that are given back to the contributors. In this case, two processes occur (which may need different contractual terms): One-way sharing between the companies and the third party, and the way back, where the roles are reversed[41].

[41] This is the case for the Urban Data Challenge Hamburg, mentioned in Section 5: The Urban Data Challenge in Hamburg: What works in practice, that served as the basis to inform the blueprint.

The proposed ontology defines a set of modules that can compose a basic architecture for its technical and organisational implementation. Two interrelated key assumptions need to be met for the ontology to work, highlighted in Figure 4.2.

- First, it is possible to generate terms governing processes at a large scale. If any data sharing scenario can consist of different processes, and each process entails a precise term, it becomes obvious that the number of terms (that can be grouped into agreements) grows very fast.
- Second, it is possible to univocally define a mapping from (i) use cases and provided data to (ii) de-identification and (mis)use risk and from there to (iii) applicable law and to data transformation procedures to comply with such laws to mitigate legal compliance risk (Section 2). This would mean that, for instance, competent data authority ruling should be deterministic given similar inputs (use cases and original data to be shared).

In the following section, we introduce tools to show how those two assumptions would work to apply the ontology to real cases, which we do in Section 4.1.4.

4.1.2. We need a complete data sharing grammar

To generate contracts at scale, we need a proper grammar to define contracts. If we can define a grammar over sharing terms, we can implement technological measures that accelerate their generation. Furthermore, we can use the grammar to recognise similar contracts and needs, and thus promote their reuse to reduce transaction costs with tested examples. As described above, defining, negotiating and implementing data sharing terms in practice is often a difficult and time-consuming task[42]. In practice, data contributors typically want to maintain as much control over data as possible. To protect business secrets that may be contained in the data, data contributors generally want to specify how the data is to be transformed and who can use the information generated for what purposes.
To have a legal basis for the provision of data, we envisage the creation of data sharing terms. These data sharing terms should contain information on who provides data, how this data is transformed, and how it is aggregated into information in a (potentially) pooled dataset. In cases where many data contributors provide various datasets for various purposes, which are used by many data users, we see a challenge in scaling up the creation of these data sharing terms.

At present, this process requires highly skilled humans, for example, lawyers (Section 2). This is very inefficient and prevents the effective use of data for the public interest. Instead, we imagine a technological system that can support the involved humans in the contract negotiation and contract creation processes[43]. This system needs to be able to express the data sharing terms to be complete and reusable across use cases. It can then propose sharing terms in a negotiation and link them to their potential (legal) consequences found in similar cases. Therefore, the negotiating actors (e.g.

[42] See Young, et al. (2019) Addressing the Challenges of Drafting Contracts for Data Collaboration.
[43] See Duisberg (2022) Legal Aspects of IDS: Data Sovereignty—What Does It Imply?

lawyers) can focus on agreeing on the chosen terms from an initial basis (not on hypothesising over their consequences), rather than starting from scratch every time.

This is possible by providing templates and rules which can (semi-)automatically create data sharing terms which are human-readable and machine-readable. A data sharing term can be a simple contract between two parties with one or a few conditions or a sum or combination of different rules and agreements of various actors. The fact that contracts are machine-readable is key to their success, because this allows indexing them, so they can be made discoverable. It also induces a similarity metric among them, so we can assess in a scalable manner how similar they are to each other and thus present interested parties with useful templates by means of recommendation techniques (e.g. with Machine Learning).

In our approach, we recommend to build on the work of the International Data Spaces Association and Gaia-X initiatives, which have been working intensively in the last few years on how data sharing use conditions can be defined, implemented, and enforced at the technical level. A more detailed description of these approaches is provided in Annex 2: Technical annexes.

4.1.3. We need a mapping between data transformations and their corresponding identification risks

In a data sharing scenario, a feedback loop happens: The intended use case determines the information needed from the original datasets provided. The information entails risk if shared. The level of risk and its associated societal costs determine in turn the applicable law(s) on each data sharing scenario. Finally, the applicable law imposes which transformations or pooling procedures are acceptable to transform the original datasets into the information needed by the data user, ideally leaving it granular enough to be fit for the original task. For a proposed grammar to be usable, we must aim at mapping these relations in an unambiguously quantifiable way, so the feedback loop can be accelerated and the legal implications of each sharing term clarified.

We acknowledge that it may be impossible to do so completely. Laws, data aggregation, and transformation procedures are binary procedures, with a yes/no-outcome (e.g. laws are either applicable or not), but risks associated with cost and probability of information disclosure[44] have a continuous range. The final decision on whether a law applies to a use case will always be carried out in natural language by a human, that is, the final balancing of identifiability probability and its societal costs (we operationalise risk as probability times costs). The tension between the discreteness of legal rulings and the continuity of risks may not always be possible to bridge in a consistent way; however, we must aspire to make the mapping as solid as possible.

In a data sharing setting, we may not know the costs in a precise way before leakage happens, but we know how to quantify the probability that it happens. In most contexts, and specifically with urban data sharing, there are two types of identification

[44] We understand information disclosure as the event that individual data points can be inferred from information even if they are intended to remain confidential. Risk is the product between the probability that an event happens with the cost that it happens.

probabilities: The chance that individual relevant business information can be isolated from the pooled datasets and that of isolating personal data from individual (customers) of the original data contributors. Because we are interested mainly in statistically relevant data, the focus should be placed on the business identification risk, even if the procedure is also useful for personal identification risk[45].

The final decision is to decide on the acceptable level of (re)identification probability and cost, and if such a level is compatible with the objectives of the use case at hand. The more information the necessary data carries, the higher this probability, ranging from fully confidential information (referring to a single individual or business) to fully general information (meant to be seen by the society at large). Costs, on the other hand, only depend on the type of information shared, not on its form.

Given a transformation, there are different ways to quantify identification probabilities (such as differential privacy[46], k-anonymity[47], l-diversity[48], t-closeness[49]), as this is a very active field of research[50]. Independent of those, the process to define the allowed data transformations we propose follows these steps (Figure 4.3):

1. Define the question that needs answering in terms of data for the proposed use case (e.g. *how many bike users this street had between 2nd and 15th August 2023*[51]), which defines the data transformations to be applied (aggregation, combination, etc.).
2. Define the scope of the identification risk: First identify the type of identification risk between business risk (i.e. *identify that a user belongs to a specific data contributor from the data*) and personal risk (i.e. *identify the personal user for each trip*). Next, assess their associated costs, e.g. *disclosure of company market share in given area* and disclosure of *relevant user location*.
3. Compute the probability of identification, given all the questions that the use case may entail.
4. Given the above, define if the probability times the perceived cost, hence the risk, is acceptable. If not, check if the level of detail of the information provided is enough for the intended use case. If it can be lowered, return to 1 and re-define the questions[52].

[45] In contexts other than (mandatory) urban data sharing, the personal risk might be more relevant, but it is not the focus of this blueprint, see Section 1.

[46] See Dwork (2006) Differential privacy. In: International colloquium on automata, languages, and programming.

[47] See Sweeney (2002) k-anonymity: A model for protecting privacy. In: International journal of uncertainty, fuzziness and knowledge-based systems.

[48] See Machanavajjhala et al. (2007) l-diversity: Privacy beyond k-anonymity. In: ACM Transactions on Knowledge Discovery from Data.

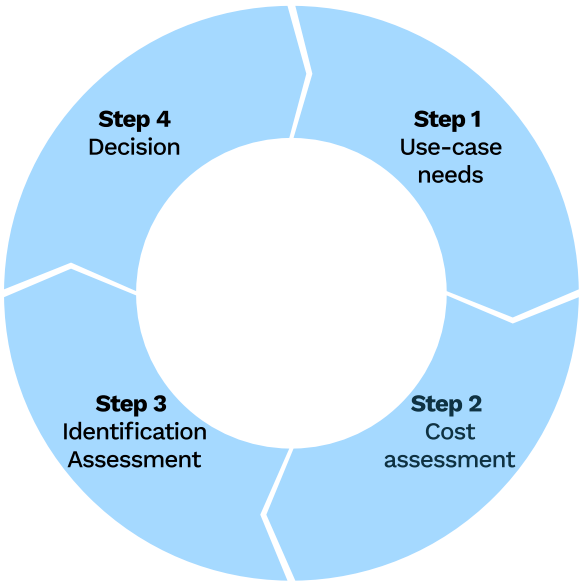
[49] See Li et al. (2006) t-closeness: Privacy beyond k-anonymity and l-diversity. In: 23rd international conference on data engineering.

[50] See Deloitte (2022) Preserving Privacy in Artificial Intelligence Applications through Anonymization of Sensitive Data.

[51] The steps are exemplified using the Hamburg Data Challenge, see Section 5: The Urban Data Challenge in Hamburg: What works in practice.

[52] In terms of differential privacy, this amounts roughly to define the *information budget*, i.e., to which barrier of probability of identification implied by the type of questions I wish to make I can attain.

Figure 4.3: Data transformation definition process



Note that the above steps may need to be iterated. In many situations, one will only find the relevant question to solve, at the relevant detailed level of information, at the end of such iterations by playing with the data.

Performing the above steps thoroughly requires specialised knowledge. However, doing them in a principled and structured way paves the road for a scalable process which should be a long-term vision. In the short term, we propose to simplify the steps in the most practical way possible: First, identify the different types of transformations involved in step 4 and categorise them (we presented a proposal in Annex 2: Technical annexes, Parametrised data transformations) and second, assign them ordered broad values of identification cost (step 2), probability (step 3) and risk (step 4) -*very low, low, moderate, high, very high*-. This initial mapping can be used already as a basis to serve the long-term enunciated objective.

4.1.4. We need a use case repository and development kit to enhance reusability across use cases

Even after performing the steps already laid out, lawyers and policymakers will still be needed for the definition of data sharing agreements. Their role is key to map the ontology to each situation, use the contract grammar, and the transformation definition procedure. Even if the above methodology is not accepted as legally binding and cannot substitute a lengthy process, the legal procedure must happen in tandem with a principled approach to achieve scaling. Reusability across cases, that can accumulate to build a coherent paradigm where to draw and automate standard procedures from, is key to scaling, as emphasised though this document in both the legal (Section 2) and governance (Section 3: How an urban data intermediary should work) sections. And for that, we presented a well detailed, practical, and usable solution for urban data sharing.

The idea of the proposed solution to achieve scalability is simple: Introduce a network of coordinated data intermediaries per city and use them as standardisation tools. The

intermediaries should define (with the help of lawyers) and enforce a common framework (the proposed ontology), to define contracts and data transforms in quantifiable ways (with the proposed grammar). They need to take special care in documenting the entire process and make it accessible to other cities (and within them, especially for the larger administrations) using unified, browsable, and discoverable templates.

Nowadays, repositories of data sharing use cases exist across the globe. See for example initiatives in France[53], in the Netherlands (CDS-M use case repository[54]) and in the United States (use case library by the Open Mobility Foundation[55]). However, these repositories[56] focus mostly on reusing good ideas, usually the use cases, for public policy, not on the legal, administrative, and organisational tools used to implement them. We strongly support that both the idea and its application should be part of the repository, and that the developed ontology and grammar are used to document the implementation. To that end, we presented a unified template to be used as a starting point to describe the different cases, and to enforce that such a template is always used. This does not mean that the template is static, but it must be versioned and developed iteratively. This, in turn, demands coordination among the data intermediaries with the joint clear foundational objective to curate such a set of standards.

The use case repository should not only be used as a tool to document the design of data sharing scenarios, but also as a tool to measure the effectiveness of their outcomes, which is a key point in public urban policies (see governance Section 3: How an urban data intermediary should work).

In Annex 1: **Use case repository examples**, we presented a proposed application of our proposed template to the three potential data sharing scenarios referenced across the sections and introduced in Section 1, that may correspond to examples present in the use case repository. To compile them, one needs to follow the steps laid out in Figure 4.4.

Figure 4.4: Use case implementation process

1. Define the use case in terms of questions to be solved, and decisions and actions that follow from the answers to these questions (who needs to take them, what do they need to know, for how long...). On the basis of that, map the questions to datasets required and to a metric to measure the success of the data sharing project to answer the questions.
2. Map the actors, datasets, and needs to the ontology. This means identifying, within the entire agreement to share data, how many processes there are and

[53] See Molins et al. (2019) Self Data cities, the playbook.
[54] Available at: https://usecasestore.cds-m.com/store
[55] Available at: https://airtable.com/app0XurVVjbou23ul/shrPf4QvORkjZmHIs/tblzFfU6fxQm5Sdhm
[56] See for instance https://oldwww.mydata.org/cases/

which roles the actors take. On the basis of the identified processes, determine which auxiliary roles are required.
3. For each process, assess the identification cost, probability, and risk and determine the data transformations accordingly. If need be, present the case to the transformation certifier to validate the law applicable.
4. Compile the data sharing terms based on the use case and constraints imposed by the type of transformation applied (observation time, scope, etc).
5. Group the atomic terms into an agreement that sets their governance (how can they be altered, expanded to new actors, etc.).
6. Document the relevant points of the discussion and the chosen terms and data transformations, link to real legal contracts and data descriptions, and evaluate the entire procedure prior to publishing.
7. If any of the tools (ontology or grammar) was not enough to describe the case and required to be adapted, issue a proposal of modification to the repository.

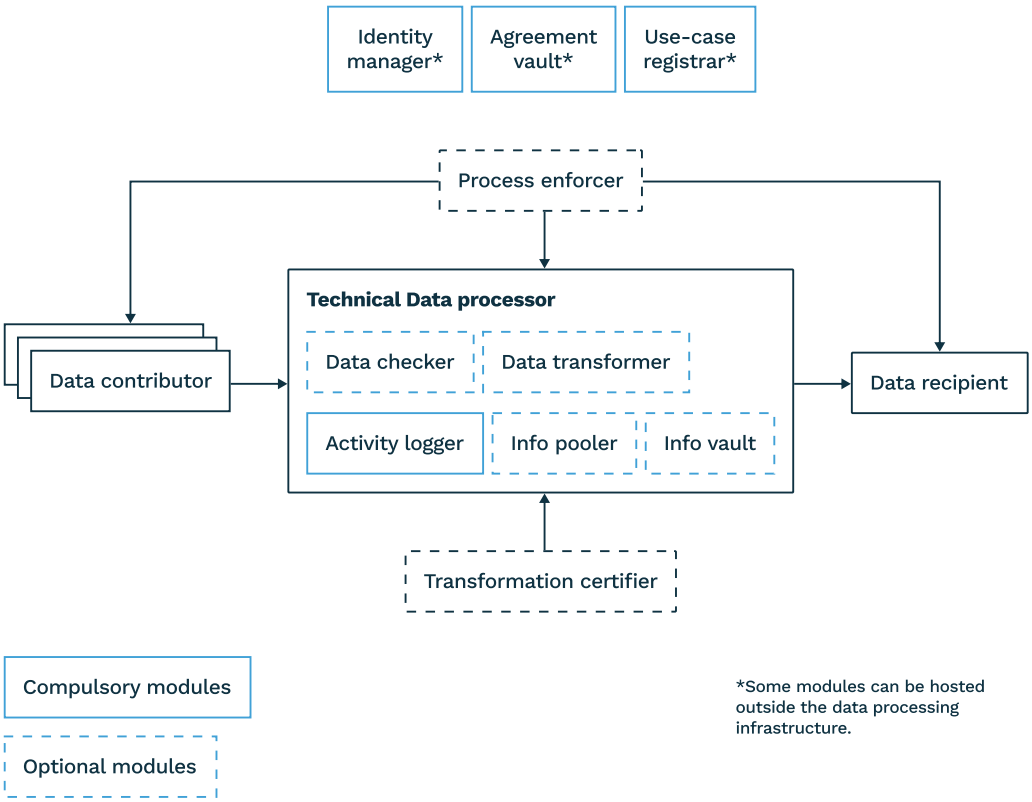
The widespread adoption of the above procedure needs to be a clear, shared goal of a network of data intermediaries to reduce the currently often prohibitive efforts for urban data sharing. Setting up the appropriate governance model and legislation to enable its existence is addressed in the other sections, but the technical implementation is complementary to the legal framework, because the former provides better capacity to enforce the latter.

4.2. How to set up the data sharing architecture

Once all the building blocks described in this section have been laid out at the conceptual level, it is time to present how they might be implemented technologically. We propose an implementation that feels like a logical derivation of all the isolated challenges and proposed solutions described in the text so far across the sections of the blueprint. With our ontology in hand, the proposed architecture is an encapsulation of the different tasks and roles present therein.

Figure 4.6 gives a high-level overview of the building blocks, which we call modules, that conform to the basic technological set proposed to support urban data sharing. Some are compulsory and others are optional, in Annex 2: Technical annexes, Detailed architecture, a detailed description of each of them is provided. We also provide a suggestion on where and under which responsibility those modules should be operated, but in some cases, this might be altered.
We leave out the description or any specific suggestion of low-level implementation and standards for those modules, because this will need to be adapted to each situation. To do so, one will need to consider aspects like technological competence, readiness, and legacy, as well as a balance between simplicity and complexity. In Annex 3: Data sharing FAQs for cities: What elements are required when?, we provide a checklist for cities to make such decisions. Moreover, we wish to emphasise that not all modules are required for all cases: Not all roles were key in the presented ontology, so many modules are also optional in this architecture.

Figure 4.5: Overview of functional modules to support data sharing scenarios



The conclude this section, we present the main principles that should guide the design and implementation of the proposed modules, beside the already presented concepts.

Principle 1: Security from traceability

Any system must be designed anticipating that anything that might go wrong, will do so at some point. Therefore, one cannot expect to find 100% assurances relying only on mechanisms set up within the system (technological). The biggest threat is always the most unpredictable, i.e., the human part. Rather than ensuring 0 level fault, we must seek processes such that audits and investigations can be put in place, and potential punitive actions activated if required. Therefore, we aim for a design that promotes and enforces rule abiding within the system and provides transparency mechanisms to trace whenever this is not true, so punitive action can take place.

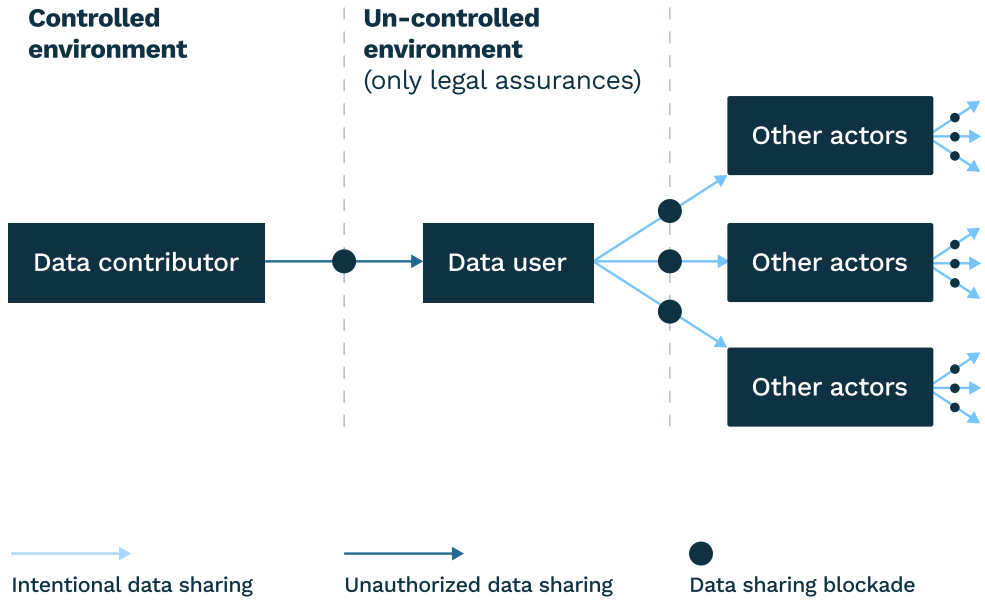
Principle 2: Irreversibility of information transfer

Actors involved in data sharing should understand that the transfer of information is an irreversible physical process that cannot be undone. This means that any effective control over information that has been disclosed[57] is lost once this happens and can only be enforced from that point onwards via legal and organisational mechanisms. Furthermore, even if one wishes to block the sharing of certain data after some time, this blockade cannot (and will not) have retroactive effects, no matter how complex and secure we aim at designing a potential solution. Such a physical limitation cannot

[57] While certain procedures such as watermarking may be applied to the support where this information is shared (i.e., the records), the information itself cannot be, as it is an intangible asset.

be overcome by technological means; hence, we argue that misuse prevention mitigation should focus on establishing the right procedures to govern data sovereignty[58] rather than aiming at complete, highly complex, and difficult to implement solutions.

Figure 4.6: Irreversibility of the data sharing process



Principle 3: Sharing among known, identifiable actors

Urban data sharing (and any type of sharing) is ultimately built on trust. Hence, identifiability of actors is a necessary ingredient for this to work under transparency and accountability principles. Any actor taking a key role in the data sharing processes must be identified and operates by acknowledging the governance practices set up by the data sharing agreements. As thus, it is responsible for all what happens within the processes and tasks that take place within their infrastructure, and for their access credentials.

Implementing such a system opens the door to the development of data usage services for (i.e. not necessarily by) the public sector. Thus, such services may be carried out by public administration themselves, public companies, or private entities, depending on the reality of each administration. However, as emphasised in Section 3: How an urban data intermediary should work, it is important to implement them following the free and open-source principle along the lines of "public money public code": We have proposed a set-up that is highly likely to be iterated and evolved, and this will most likely not be possible with closed source implementations. Furthermore, the principles of *transparency and accountability* would hardly be respected. Moreover, an open-source implementation guarantees that small cities may access the same advancements that larger administrations implement (and fund), and this in turn may

[58] We follow the definition of data sovereignty as the process by which authorised actors, with defined use cases, temporal, and spatial limits access determined information; which is tracked, enforced and audited, and backed by relevant legal, technical, and organisational procedures.

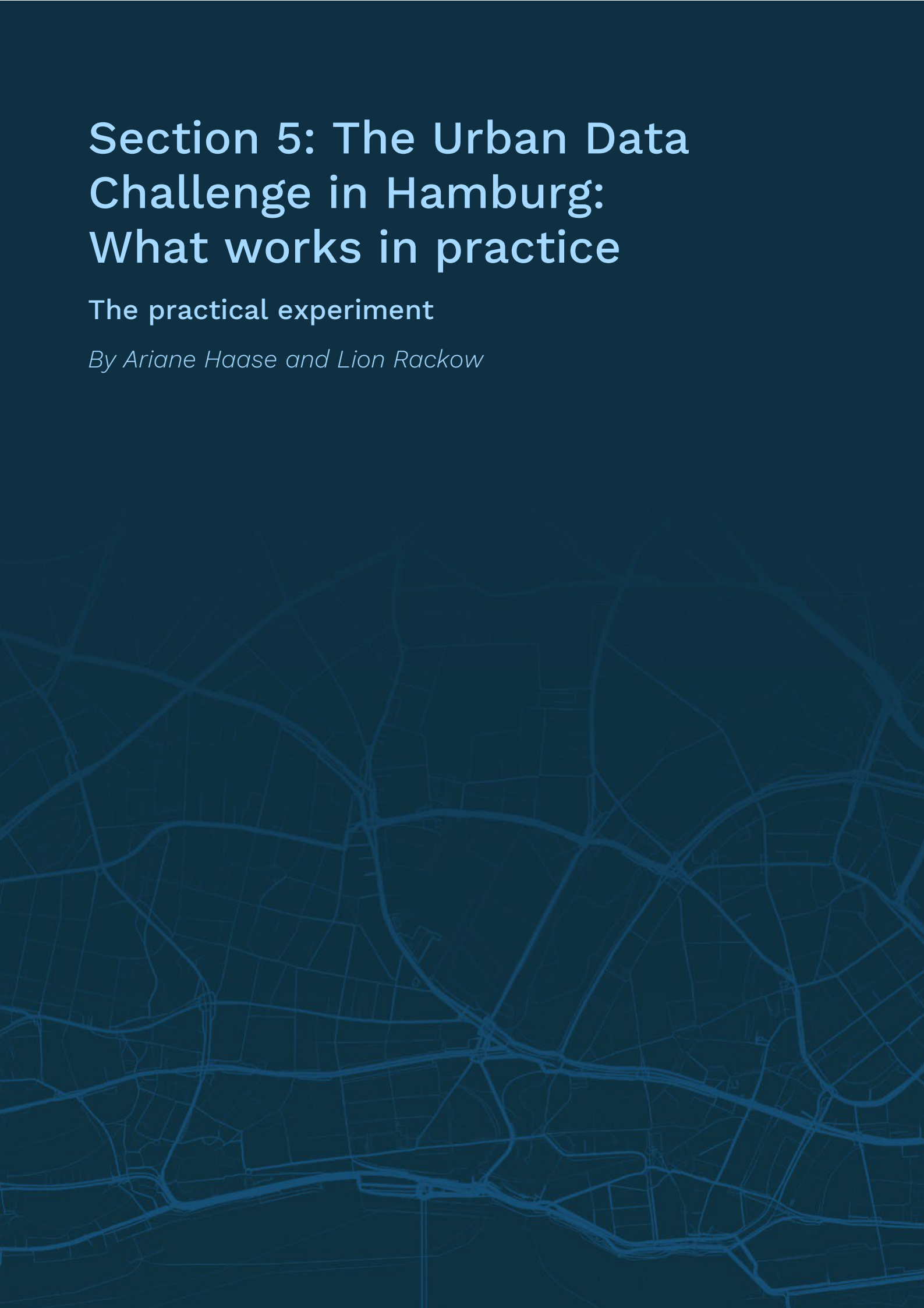
help the establishment of an ecosystem of SMEs that may support such small cities in their implementation, and standardisation, efforts.

In Annex 3: Data sharing FAQs for cities: What elements are required when?, we provide a *frequently asked questions list* (FAQ) specifically meant as a guide for administrations wishing to implement the proposed tools present in this document.

Section 5: The Urban Data Challenge in Hamburg: What works in practice

The practical experiment

By Ariane Haase and Lion Rackow



The Urban Data Challenge Hamburg is an innovation experiment conducted by the City of Hamburg and The New Institute as part of the joint project The New Hanse in 2022 and 2023. The objectives of the experiment were twofold: First, the challenge served as practical testing ground for the selection, creation, and evaluation of a feasible data sharing governance framework for public-private relations in the City of Hamburg and was therefore essential for the development of this blueprint. Second, the Urban Data Challenge was run to trigger the development and prototyping of a data-driven solution to a real challenge of Hamburg's administration in the field of micromobility and green infrastructure planning, through an open collaboration process. The guiding question for this endeavour was:

"How can we gain insights into cycling & micromobility flows and the associated ecosystem in Hamburg to make the city more liveable and sustainable?"

The following describes the genesis, difficulties, and lessons learned from the Urban Data Challenge. With regard to the overarching purpose of this blueprint, the focus will be on practical insights on how urban data sharing initiatives can be feasible on the ground, as well as what is holding them back.

5.1. What an urban challenge is

An urban challenge is a prize competition that aims to address an identified specific societal, environmental, or technological need of a city and is open to the public including the innovation ecosystem.

This method has proven successful in stimulating innovation by harnessing the appeal and excitement of competition and prize money[59], while addressing a public need that benefits citizens. An illustrative example of this is the City of Barcelona[60], where the challenge format has been applied across various sectors (e.g. mobility, tourism, technology) and proven to be very successful over the course of many years. Specific examples are challenges on the reduction of the environmental footprint of the technology sector and on ensuring the proper use of reserved parking places.

When comparing the urban challenge approach with traditional procurement processes, the following benefits can be perceived:

- Serves as a testing ground of innovative solutions reducing the risks of large-scale deployments of solutions that might not work.
- Lowers the entry barriers for (small) companies and diverse providers to take part in urban service delivery, answering to problems identified by the city.
- Fosters collaboration across city departments and among the innovation and e.g., data & mobility ecosystems from the public, private, academic sector, and civil society.

[59] It is essential to ensure a low barrier initial proposal process and to financially compensate the (final) participants for their effort as well as compensate the winner in the execution phase.
[60] https://ajuntament.barcelona.cat/agenda2030/en/projects/innovaccio-2030/challenges

- Builds trust in the local government, as it sends out a message of openness and collaboration.
- Enables the city to deliver better quality services, thereby improving their citizens' quality of life, while also improving the policy frameworks for instance if the challenge involves data governance.

Usually, urban challenges are also part of a broader approach about accessing and sharing data, digital infrastructures, solutions, and learnings from past experiences whether successful or unsuccessful.
Sharing resources and partnership building is key in innovation because it shows a better way to invest public resources both economic and human resources by learning from what works and what does not in similar situations, and by integrating new perspectives, participation of stakeholders, and talent into the decision-making process. Solutions that work should be shared and then adapted to the local context, and eventually scaled after a process of evaluation and allocation of the required resources (for example via public procurement or public financing).

Building on the principle of sharing the challenge approach is generally guided by the values of:

- **Participation**, meaning that the city representatives are willing to understand and listen to citizens' and other stakeholders' needs and integrate these needs in the decision-making process (to pair them with what the market can offer).
- **Openness** by sending the message that the city is inviting different stakeholders to propose and co-design solutions, rather than purely making assumptions about areas of improvements and how to manage them.
- **Partnership**, as it fosters internal and external cross-sector collaboration. Internal collaboration means communicating and working together across several departments, for example, data and mobility. External collaboration means forging partnerships with the private sector (from large companies to SMEs and start-ups) as well as bringing in the social (civil society organisations, NGOs) and the academic sectors (universities and research).

5.2. Why Hamburg is the ideal place to run an urban challenge

In Germany, the City of Hamburg is the perfect laboratory to run an urban data challenge focused on sustainability, with its progressive digital policy goals[61], its net-zero agenda and its unique transparency law that gives the citizens comprehensive rights to information as well as its advanced technical infrastructure and capabilities with the Urban Data Platform.[62] Building on these goals and capabilities, the City of

[61] ITS Strategy for Hamburg, https://www.hamburg.de/contentblob/5934418/2afc89cd64f950803e25689ad3e5db87/data/its-strategie-fuer-hamburg.pdf

[62] As a central platform for Hamburg, the UDP pools urban data and makes it available to users – interoperable and independent of the manufacturer. It connects IT systems from a range of categories such as business, science, culture, environment, and transport. Users can configure the data according to their individual requirements and evaluate, analyse, and retrieve data in real time. Public administration, civil society, science, and the business community benefit from the synergies and value added. The UDP was developed by the

Hamburg set very ambitious goals when it comes to the transition to green mobility: By the end of the decade, the share of all journeys made by public transport (including car sharing), cycling trips and foot traffic shall be increased to 80% of the modal split, with the specific target for bicycle traffic being an increase to 25%–30%[63].

Accordingly, one of the priorities of Hamburg's administration is to improve the quality of cycling in the city, for example by experimenting with temporary pop-up bike lanes to see what effect they might have on mobility flows. The practical learnings gained from these measures (supported by qualitative data from interviews) will help the administration to assess the potential impact of infrastructure changes on the modal split in the city and will also provide an evidence base for deciding whether or not to roll out pop-up bike lanes more widely, ultimately make them permanent and other infrastructure adjustments.

At present, however, the City of Hamburg lacks representative data, such as on the distribution of cyclists on Hamburg's road and path network, their average speed and origin-destination relationships. This makes it difficult to evaluate the impact of infrastructure measures. In the past, the Authority for Traffic and Mobility Transition has therefore commissioned manual counts of how busy a street or intersection is to assess the effectiveness of a pop-up bike lane[64]. A key obstacle to a more data-driven approach is that data on cyclists is available to the city only in a very scattered way from various providers and with certain biases, and there is currently no consolidation as there is for car traffic (e.g. from Inrix, TomTom). To develop and evaluate effective infrastructure measures and to allocate public space according to the real needs of citizens, it is therefore crucial to create a solid database and specific digital tools for the city that help with visualising and analysing non-motorised mobility flows.

With regard to the other objective of the Urban Data Challenge, which is to test new methods of data sharing and data governance for the public interest as developed and proposed in this blueprint, this concrete challenge of the City of Hamburg provides a very interesting use case.
This is because, first, the planning problems associated with the rapid growth of shared micromobility services are ubiquitous in all major European cities, and, as argued above, data is a key component in regulating markets and finding solutions or delivering services for many of these challenges.
Second, the question of how exactly this data, which is mainly generated by private providers of shared micromobility services, could be made available to city administrations and the innovation ecosystem at large, is linked to issues of leveraging

Agency for Geoinformation and Surveying (LGV) with the aim of making data usable for various applications and processes, including participation processes in urban development, intelligent traffic management, and social infrastructure planning.
[63] Coalition Contract/ Koalitionsvertrag über die Zusammenarbeit in der 22. Legislaturperiode der Hamburgischen Bürgerschaft zwischen der SPD, Landesorganisation Hamburg und Bündnis 90/Die Grünen, Landesverband Hamburg, https://www.hamburg.de/senatsthemen/koalitionsvertrag/verkehr/#marker02
[64] See for example the evaluation report of a different pop-up bike lane in the Hallerstraße, which was published June 2022: https://www.hamburg.de/contentblob/16328090/6b4b783311ea31cbe58024ac781eded6/data/endabgabe-hallerstrasse.pdf

urban data for the public interest, while preserving data protection and balancing public and private interests.
Therefore, it brings into focus the key challenges of governing urban data and digital infrastructures, while putting citizens and the public interest first.

To explore new data sharing methods in practice, and in light of the success of urban challenges in fostering innovation in Barcelona and other cities, the project team from The New Institute and the City of Hamburg therefore decided to focus the Urban Data Challenge on the issue of micromobility in Hamburg. The underlying assumption was that by solving a concrete problem in this field through the use of shared private and public data, the potential of data sharing in the public interest for the sustainable transformation of cities could be demonstrated.

5.3. How the Urban Data Challenge was implemented

To design and implement the challenge format in Hamburg, The New Institute drew upon the knowledge of experienced external consultants, directed by Francesca Bria, who have been previously working with the challenge format in Barcelona for many years. The City of Hamburg itself had no previous experience with running a challenge, but exchanged knowledge with other German cities, such as Leipzig[65], which had already worked successfully with smart city challenges.

Overall, the challenge design process can roughly be outlined as follows:

Table 5.1: Conceptualisation phase of Urban Data Challenge Hamburg

Conceptualisation phase	
Q4 2021	Initial agreement between City of Hamburg and The New Institute on project and an urban challenge as experimentation format
Q1 & Q2 2022	Workshops with different stakeholders/departments in Hamburg's city administration to identify the challenge. Several rounds were required, as the city did not have a pre-existing overview/catalogue of specific use cases where private data was required to solve a problem and trigger sustainable innovation.
Q2 & Q3 2022	Workshops and feedback rounds to shape precise challenge question from city internal view and discuss potential data sharing frameworks
Q3 2022	Signature of official Letter of Intent, outlining project details
Q4 2022	Workshops and feedback rounds with six external micromobility companies/organisations to verify and advance precise challenge questions and identify potential data partners
Q4 2022	Decision on procurement format/procedures for Urban Data Challenge Hamburg and finalisation of challenge definition
Q1 2023	Various feedback loops for procurement documents to be published end of February 2023

At the end of the conceptualisation phase that also involved experts from the Authority for Traffic and Mobility Transition the final challenge question was the following:

[65] https://digitalcampus.leipzig.de/sccl/sccl-2022/

"How can we gain insights into cycling & micromobility flows and the associated ecosystem in Hamburg to make the city more liveable and sustainable?"[66]

To enable a data-driven investigation of this question, two types of data have been included: a comprehensive set of exclusive private micromobility data was made available alongside data which the City of Hamburg already makes publicly available on its Urban Data Platform. The private datasets were provided by the partner companies Bolt[67] and IoT Venture[68], who, together with 11 other organisations active in the field of micromobility in Hamburg, were invited to become partners in the Urban Data Challenge Hamburg. Although the other organisations did not participate for various reasons, including lack of resources as well as bad experiences in working with cities in the past, these two companies decided to voluntarily share data including a range of movement and vehicle data for a given time period. Their reasons to support the project with their data was mainly their interest in the success of the Urban Data Challenge, i.e. better evidence to support political intervention and infrastructure planning for more micromobility as well as close collaboration with government and administration, which would ultimately help their business. As part of the Urban Data Challenge, these datasets are to be analysed, combined and put into relation (optionally with other datasets proposed by the participants). The goal was to develop a holistic view of micromobility behaviour and, if possible, to enable various analyses. The most important use case to be considered is the change in mobility behaviour on the Reeperbahn due to the new eastbound pop-up bike lane from November 2022. By means of a before-and-after analysis, the city wants to investigate the resulting effects in the neighbouring districts of St. Pauli, Altona-Nord, Altona-Altstadt, Neustadt, and Sternschanze and thereby derive results for further projects. Beyond the "Reeperbahn" use case, the expected analysis results are to be used to implement various infrastructural and traffic improvements.

5.4. What we learned about data sharing in practice

The Urban Data Challenge, as a collaborative experiment with the City of Hamburg, provided a unique opportunity to gain insights into the city's approach to data governance. Throughout the project, it became evident that a robust commitment exists at the executive city level to confront challenges and seize opportunities linked to urban data (see Section 1). This was also reflected in the substantial support for The New Hanse project, allowing for a collaborative and hands-on exploration of novel approaches to governing and sharing data within the City of Hamburg and beyond.
However, at the operational level, engaging in an innovative data project proved to be challenging, albeit ultimately successful. A key reason for this was that data governance is a cross-departmental issue that requires mediation among various city stakeholders

[66] To allow for a more holistic analysis of the use of the bike lane, the term micromobility was specifically added to broaden the scope of the challenge to include (data from) all vehicle types allowed on bike lanes and also the pop-up bike lane in question, such as (electric) cargo bikes and electric scooters. The term refers to small, lightweight vehicles (such as bicycles, e-scooters, cargo bikes, e-mopeds) that typically travel at speeds below 25 km/h and can be either human-powered or electric.
[67] https://bolt.eu/de-de/
[68] https://iot-venture.com/

with sometimes divergent perspectives. A common vision, clear roles and responsibilities among the different stakeholders and regular check-ins at the working level were essential to navigate this complexity and propel the project forward. Another crucial lesson derived from the practical experiment is the substantial interest within the broader urban innovation ecosystem in initiatives such as the Urban Data Challenge, showcasing a desire to access new (private) urban datasets to develop digital tools and solutions addressing the many challenges faced by cities today.

In consideration of these insights, a promising window of opportunity emerges for designing and implementing projects that envision innovative urban data governance models for the public interest. Although this prospect is exciting, it is crucial to approach new endeavours with an awareness of the specific challenges and pitfalls that exist.

Through the implementation of the Urban Data Challenge and interactions with the data contributors and in particular with Hamburg's Data Protection Agency (DPA) around the Hamburg Commissioner for Data Protection and Freedom of Information,[69] The New Hanse project has gathered valuable insights into the practical challenges of urban data sharing. The main points of discussion included: i) the analysis of partner companies' privacy notices, ii) the role the City of Hamburg shall play in the set-up of the challenge process and in the transfer and governance of data and iii) Discussion of (pseudo-)anonymisation-utility trade of. All the insights gained in these exchanges were highly pertinent to the legal, the governance, and the technical aspects of the Urban Data Challenge and The New Hanse project at large.

The key takeaways are listed below and will also resonate with some of the points and arguments in the other sections of this blueprint, as they have served as important practical examples and starting points for discussions within the work streams of the Data Commons Working Group:

1) **Expect the unexpected in an emerging field**
   The first learning was – what else could one expect from a data-driven innovation project that is a new field of action– that one should always expect the unexpected. In this case, the unexpected was a data protection incident from another European Member State. In this reference case from France[70], the sharing company CityScoot was accused of having collected geodata with allegedly inappropriate frequency and detail. On the basis of this case, the Hamburg DPA raised concerns regarding Bolt's geodata provided and outlined possible implications for the City of Hamburg, when getting involved with such data. Learning in practice, with real cases examples in a field that is rapidly evolving but still emergent, is essential to come to mature policies, and practical solutions.

2) **The city as a facilitator (or not) - part I: the Urban Data Platform (UDP) could not serve as a technical data handler**

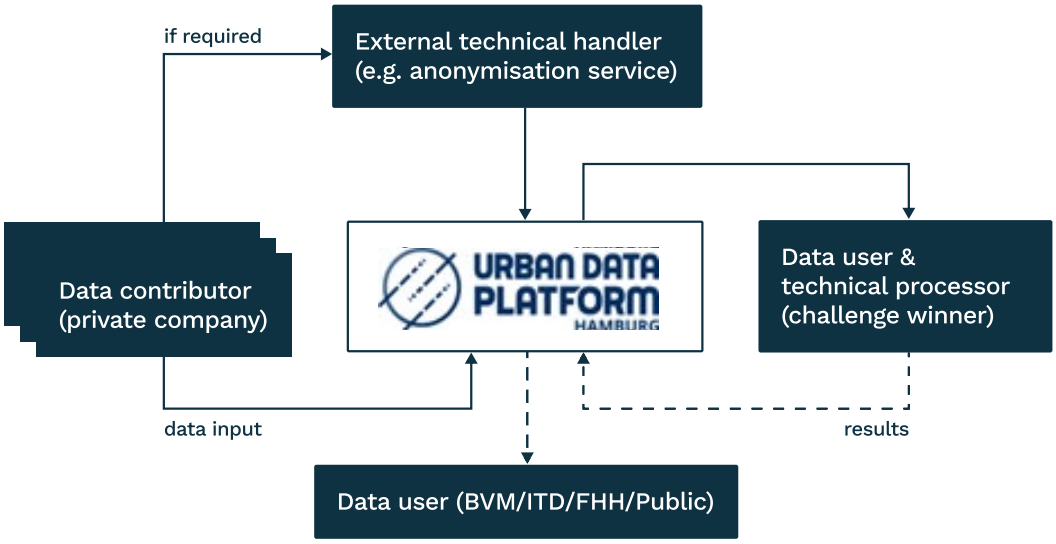
[69] Thomas Fuchs, Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit (HmbBfDI)

[70] https://www.cnil.fr/fr/geolocalisation-de-scooters-de-location-sanction-de-125-000-euros-lencontre-de-cityscoot

To enable the transfer of data to the challenge winner, the data user, for the development of a data-driven solution to the Urban Data Challenge, a specific technical and legal set-up had to be developed that takes into account data protection and trade business secrets constraints, while ideally also testing new methods of urban data sharing (Please also see the technical Section 3 for a more detailed introduction to the ontology)

In an initial legal assessment[71] for the Urban Data Challenge Hamburg, Max von Grafenstein argued that for a long-term vision the Urban Data Platform (UDP) should not act as such a data intermediary, due to its lack of independence from the City of Hamburg. However, it was planned to include the UDP at least as a technical data handler, at least for this one-off challenge, making use of the already existing capacities and competencies with regard to data availability, processing, etc. The idea was that the UDP would not only provide the aforementioned public city data for the challenge, but also act as a technical data handler platform for the data sharing process between the partner companies and the challenge winner.

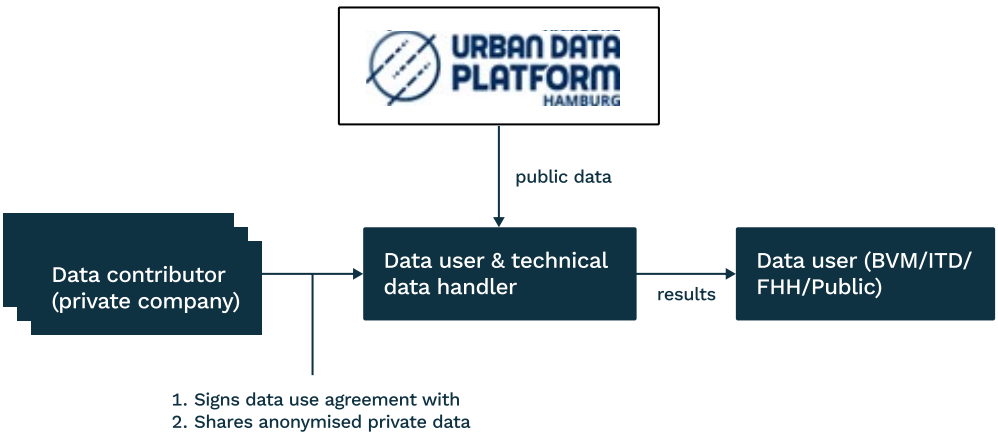
Figure 5.1: Envisioned techno-legal set-up of Urban Data Challenge



However, the complex discussion with the DPA about the possible role of the city as a joint controller and the resulting legal problems due to the data protection incident in France led to the decision to keep the City of Hamburg, and thus the UDP, out of the data sharing processes as far as possible, also with regard to technical issues.
This had the consequence that the data could not be shared via the UDP, but instead had to be exchanged directly between data contributors Bolt and IoT Venture and the challenge winner (please see Figure 5.2 with the finally implemented techno-legal set-up).

[71] Available under: https://thenewhanse.eu/en/a-first-legal-assessment

Figure 5.2: Final techno-legal set-up of Urban Data Challenge Hamburg



We think this was a practical solution to execute the role of the data handler of the challenge in the given framework, but that the UDP might be able to perform this function in the future with a proper assessment of the use case scenario.

3) The trade-off between data anonymisation and the utility of data

Regarding the topic of anonymisation, the original plan for the Urban Data Challenge was that only the data from IoT Venture would need to be anonymised before being shared with the UDC winner (because IoT Venture has promised this to its end users in its initial privacy policy), while Bolt's data could be transferred as raw data. This would have given the data user, i.e. the winner of the UDC, more leeway in using the data provided by Bolt. The reason for this is usually described with the so-called anonymisation-utility-trade-off: The more the data is anonymised, for example through aggregation of the data or adding noise (i.e. untrue information) to the data, the more of the information originally contained in the data is lost and the less a data user may use it for his own analysis purposes. However, due to the reference case that happened in France, the data protection authority recommended that the data from Bolt should also be anonymised before transfer. Even though this limited the utility of the data, the data users have still been able to achieve their analytical goals.

4) The city as a facilitator (or not) - part II: the limitations of the city in giving legal advice

The discussions with the DPA also had an impact on the legal set up of the data sharing process for the Urban Data Challenge. Initially, the idea was to use a Joint Controllership Agreement (JCA), which is a multi-party legal agreement that specifies legal rights and obligations of the data contributors (partner companies), the data user (challenge winner) and the role of the City of Hamburg. However, this idea was discarded for two reasons: First, due to the decision of the City of Hamburg to not become a (joint) controller of the data (see 2.). Second, because of the anonymisation of all data prior to the transfer (see 3.), which effectively meant that the data contributors and the data users did not become (joint) controllers either. Consequently, there was no longer any need for a JCA.

Notably, however, a consequential problem arose from the specific requirement that all data be anonymised prior to the exchange (see point 3 above). The reason for this was that, from the DPA's point of view, this requirement raised the question of whether the city was setting the purpose of anonymisation and was thus once again - unintentionally - becoming a (joint) controller. Consequently, the city refrained from providing further support to the data contributors to avoid becoming responsible under data protection law.
Because of the strict view of the DPA, the city had to take on the data protection responsibility for this advice (i.e. become a joint controller) as soon as it wanted to help with the sharing of data through well-intentioned advice alone. Because the city did not want to take on any responsibility for Bolt's data in view of the data protection incident in France, it was no longer allowed to help with sharing at all. The deterrent effect that any assistance, no matter how small, leads to legal liability is likely to be transferable to other use cases. This is one of the main reasons why a data intermediary that not only provides assistance but also bears the legal responsibility for it is promising for the success of data sharing in practice.

5) Enabling data-driven innovation while safeguarding data protection

The process of finding a functioning legal set up for the transfer of data in the Urban Data Challenge was fairly complex: A group of several lawyers and experts from the city (data protection lawyer from Hamburg's Authority for Transport and Mobility Transition, lawyer from the Senate Chancellery) and an external law firm, with experience in the intersection of data protection, business secrets, and contracts as a sparring partner, as well as legal expertise via The New Institute through Max von Grafenstein (lawyer, researcher and author of the legal Section 2 of this document) were involved in the discussions about the requirements and features of such a set up. The issues they had to resolve in this context were not trivial and required different perspectives, professional debate, and much perseverance. In essence, the challenge was to find the right balance between enabling data-driven innovation through data sharing and safeguarding key data protection principles.
In this context, it was important to involve the Hamburg Commissioner for Data Protection and Freedom of Information, in the jury of the Urban Data Challenge[72], which was tasked with selecting the final two participants from the top five on Pitch Day on 5th May 2023. In line with his expertise, the DPA specifically evaluated the data protection concepts of the proposals, as it was important for the success of the challenge to have the support of the DPA and to start a discussion and exchange of views on the proposed data intermediary setting.

6) Difficulties in encouraging voluntary data sharing

As described above, a unique feature of the Urban Data Challenge Hamburg was that it integrated both private and public data relevant to the data governance

[72] Full jury consisted of: Thomas Fuchs (Hamburg Commissioner for Data Protection and Freedom of Information) Christian Pfromm (CDO of the Free and Hanseatic City of Hamburg), Prof. Dr. Jana Kühl (HAW Ostfalia), Edda Becker (Innovation Expert and Mentor for Startup Teens), Diether Schönfelder (CDO of the Authority of Transport and Mobility Transition), Aline Blankertz (Wikimedia), Mario Schmitz (IOT Venture), and Natascha Spörle (Bolt)

question. To facilitate this, various companies and other organisations active in the field of micromobility in Hamburg were interviewed by the project team, invited to workshops, and made aware of the project and its goals. In the end, 13 of these organisations were contacted and invited to become data partners (contributors) for the challenge by donating data and thereby supporting the green mobility transition of the City of Hamburg. However, only two companies agreed to become partners and voluntarily provided their data. Reasons given by the other organisations for not participating included lack of resources, lack of interest in the broader topic of B2G data sharing, existing involvement in other initiatives exploring the issue as well as poor experiences of working with cities in the past. Notably, the legal challenges mentioned above, which arose during the challenge, were not even an issue at the time of recruiting the partner companies. If these had been anticipated and clearly communicated to potential partners at the beginning of the process, it is likely that even fewer interested parties would have been found to contribute their data.

In summary, the practice of urban data sharing is fraught with many complex challenges and impediments. At the same time, however, there is a discernible shift among municipal stakeholders towards recognising the imperative for change and harnessing the transformative potential of data for the public good. A practical illustration of what can be achieved in this way is the Urban Data Challenge which stands as a positive example of leveraging a mix of private and public data in the public interest. Soon, the forthcoming insights will provide the City of Hamburg with a better understanding of how green infrastructure is used and what effect it may have on the overall mobility transition.

We therefore hope that this case serves as an inspiration for other cities to emulate, drawing from our shared experiences and ideas: This overall blueprint provides the necessary tools to overcome the challenges and blockages on the way and to even go beyond individual cases and towards a more progressive approach to urban data governance as a whole. As articulated earlier (refer to Section 1 and Section 2), a crucial first step would involve cities leveraging their legislative authority to mandate data holders to share their data. This action then lays the foundation for new data-driven innovations and evidence-based policy measures for the public interest.

Annexes

Annex 1: Use case repository examples

In this section, we provide four examples of *use case cards* that may compose building blocks of our proposed use case repository. As explained, the focus here is not only to share examples of good applications of urban data sharing, but also to make available the organisational, technical, and legal tools used to attain them.

We provide first a template on what a use case card should contain at least, and then develop them for the four example cases described in the introductory Section 1 and governance Section 3. Out of the four developed cases, the one with the greatest detail is the only one that has been implemented in practice in the Urban Data Challenge in Hamburg (see Section 5 in the blueprint). However, these cards remain on a provisional basis given that the exact framework for adequately describing the use cases still needs further investigation. Even more so, the other cases may be considered potential sketches and drafts of what the final cards should look like.

Use case template card

Precisely setting the content and form of the cards present in the repository lies beyond the scope of this blueprint, as it involves the definition of many nuanced details. We nevertheless presented a basic proposal meant as a starting point for its iteration.

The first element described should be the ***use case*** motivating data sharing. It should be stated as a question (or group of questions), as precise as possible, and its derived actions to be taken once it is answered. Thus, the description of the use case illustrates the value of the data (sharing and processing).

The second element should describe the **datasets provided** by the data contributors as well as in which context the data has been collected by whom for what initial purpose(s).

The third element is the description of the potential risks of the data sharing and the **limitations involved in each process** of the data sharing to control and mitigate the risks. These include limitations on all three data governance layers, i.e. the regulatory, organisational, and technical layers. These limitations describe, for example, who is allowed to use the data, for which time, for which concrete purposes. Also their information scope limitations, i.e., what the other actors are not allowed to infer from the original data provided.

The final element present should be a detailed description of all the data sharing processes to be implemented to exploit the value of the data and control the risks, in the best case, at minimum cost. For each process, an **ontology mapping** that describes all the actors present in the data sharing set-up, their roles, and tasks should be attached. This includes, for each process, a description of the data transformations applied, a quantification of their identification probability, cost, and risk, their law of

application and the potential certification by the transformation certifier. Finally, a description of the data sharing agreement contract involved should also be attached.

B2G2S: Urban Data Challenge Hamburg case

Use case: Analysis of changes in mobility patterns due to interventions on the street (adding a pop-up bike lane). The city does not have granular enough information to infer such changes and wants to understand its effect to prioritise planning of further bike lanes across the city by using a third party with analytical expertise to do so. The third party is nominated after a public challenge (competition of ideas) has taken place.

Datasets: IoT company of on-board bike devices readings of time stamped bike trajectories (speed, acceleration, coordinates) identified by pseudo via hashing device ID. Trajectory data from on-demand scooter trips (origin, destination, coordinates, speed...) identified by pseudo via hashing vehicle and user ID. Both datasets cover three periods within a time span of 2 years. The data has originally been collected by the companies from their end users when using the bikes resp. kick-scooters on the basis of a contract or the legitimate interests-clause (Art. 6 sect. 1 lit. a) and f) GDPR). The companies transferred the data on a voluntary basis to third parties who only used the data for statistical purposes. However, the companies already anonymised the data by aggregation before transferring the data.

Limitations: Data contributors should not provide personal data but only anonymised data and should not see the other data contributors' records for reasons of trade secret protection. Final results (both public and private) should not allow discerning individual behaviour from data contributor's users. Final results from the study could generate two versions of the analysis: An aggregated one to be published as open data, and optionally, a detailed one for the data contributors and the city (if applicable).

The scope of the agreement over the data is for a one-shot study, so no additional actors may join and terms cannot be changed. The public analysis is freely published as open data without restriction of use.

Ontology map: The data sharing agreement is composed of two processes. The actors present in it are: The city administration (FHH), the Urban Data Platform (UDP), the challenge winner (CH) and the companies contributing data (IoT and Bolt). Along the process, the data intermediary (in this particular case, the FHH and the New Institute - TNI) only takes an enabler role by recommending contracts among the parties. There is no process enforcer, and the transformation certifier is the Hamburg Data Protection Authority.

Process one In this process, the two companies act as data contributors, the challenge winner is the data user, and there is no technical data handler involved as the exchange is direct and *as-is*. The challenge winner implements pooling transformations by delegation from the data contributors, because it will combine datasets from both companies providing data. The pseudo-anonymisation is performed individually by each data contributor prior to the start of the process.

Process two (A and B): In this process, the two companies and the city (A) and the general public (B) are data users, while the data contributor is the challenge winner. The data contributor is the one implementing aggregation techniques to make the results ready for both the private (to A) and public (to B) delivery by delegation from

the other parties. The UDP (the city infrastructure) acts as a technical data handler that hosts the final results to be shared with the public as open data on behalf of the FHH.

Law of application The data transformations are imposed and validated by the transformation certifier on the grounds of individual data protection measures for citizens using the commercial services of the data contributors. Business trade secrets apply to process 1 and 2A, and the data is confidential among the data contributors and users. The Hamburg transparency law applies to the version of the study received by the city which is made available as open data, process 2B.

Identification risks We deal with a case of voluntary data sharing by companies that do not have competing interests and whose names are publicly available. Thus, from a business identification perspective, the identification probability is very high, while its associated cost is very low and hence the overall risk is low. On the personal identification side, the probability of identification of individuals from the data is very low, and its associated cost is low (the use of scooters is not recurrent and would display trajectories that are not frequent in general).

Data transformations On the business protection side, no data transformations are required.
For process one, on the personal protection side, two individual transformations of randomisation are imposed by the transformation certifier to prevent personal identification of trips from the Bolt dataset. The trips corresponding to the same user via a user ID are assigned randomised (but unique) user IDs in blocks of consecutive days, to prevent identification by recurrent patterns. Moreover, the start and end location of each trip are randomised within a certain radius.

For process two, the transformation procedures have not been defined yet, as the results from the analysis generated by the data contributor (Challenge winner) are yet to be produced. However, it is to be expected that an aggregation procedure will be required to avoid exposing individual behaviours to the open-data version of the results due to both personal and most importantly business protection reasons (process 2B).

Agreement contract The entire data sharing has been compiled under the form of two contracts between each of the companies providing the data and the challenge winner on the one hand for process 1, and another contract between the FHH and the challenge winner for process 2.

B2G: Mandatory urban data sharing via licences

Use case: The city imposes that as part of the authorisation to provide services in the city, and the companies have to share the generated data with the city. The city can use the information to enhance legislation, validate proper use of urban space and better planning. Optionally, in return, the data intermediary nominated by the city can provide additional services to the companies thanks to its central and neutral role (see other cases in this repository). In this example, we use shared micro mobility operators, because this is an area where city governments have been proactive in establishing data sharing mandates, but it can be easily - and has in fact already - extended to other areas such as short-term accommodations or last-mile delivery.[73] For a well-

[73] See Stephen Larrick (2022) cited above.

developed example, see the data sharing standards and use cases developed by a collaboration of cities in the Netherlands[74].

Datasets: All trips generated in the city in real time (once they are finished) and their trip statistical details (average speed, trajectory etc) are provided to the data intermediary. The city has the right to generate specific questions on a scope approved by the data agreement to be solved by the data intermediary (*i.e. tell me how many trips were generated from A to B on the 3rd week of the year*) and define enforcement alarms (*i.e. notify me whenever a trip ends off bounds*/exceeds allowed speed, etc), otherwise, it does not have access to granular data.

Limitations: The granular data is strictly confidential to the data intermediary. The scope of allowed questions by the city must be in the agreement, if new scopes for questions appear, those must be approved via appropriate governance mechanisms. The data should never include information about customer ID or personal data related to them.

Ontology map: The data agreement is composed by a single process. The data contributors are the licensed companies, while the data user is always the city. Being a public enforceable process, a process enforcer (public agency) can be nominated by the parties and the data protection authorities must act as transformation certifiers.

Process one The process is simple, data contributors stream data as it is produced to the data intermediary/technical data handler if it is delegated to do so. Then, if alarms are triggered, information is sent to the city for enforcement (data user). On an on-demand or periodical basis, the data contributor generates a report that answers the questions of the city that are within the accepted scope. For questions outside of the scope, a governance process must be defined to either accept them and include them in the atomic contracts, or reject them.

Law of application A detailed analysis should determine if transparency laws apply to this case. Data protection law should play a minor role as client individual details are never explicitly shared in the data (so the most data one can isolate from an individual is a single trip). Business trade protection and public licensing law applies in this case.

Identification risks The personal identification probability is very low. Identification cost can be high for services of recurrent use by users (energy for instance), but for the particular case of non-recurrent mobility can be considered low, so overall risk is low. On the business side, because data is only shared with the city, one can consider identification probability very high but, given its *neutral* role, costs are low so overall risk is moderate[75].

Data transformations In this case, the data transformations will be of type *pooling* from the different companies and most likely *aggregation* only, without the need to use other, more detailed operations, due to the low business identification risk involved.

Agreement contract The agreement contract in this case lies implicitly within the licence provided by the city to the service providers. Each data contributor subscribes to an individual agreement with the city that should include the same bindings in each case imposed to the service provider to avoid differential treatment. To summarise, it is the reuse of a single contract many times with the different data contributors.

[74] Available at: https://cds-m.com/

[75] Special care should be taken on preventing leaks, but this is outside the scope of this category and is a general principle.

B2G2B: Aggregated market data [hypothetical]

Use case: Companies within the same market and competing interests want to obtain a clear picture of the market growth and penetration, to prioritise business decisions related to marketing investment per area or per product, however, they need an external actor to pool the data for them. This applies to many examples, for instance the pharmaceutical industry. For the urban data case, this is relevant for on-demand micro mobility services, for instance. Another similar case which establishes data sharing obligations to private actors, which is then aggregated and analysed by public entities to increase efficiency and promote innovation, is the case of Chinese electric vehicle manufacturers.[76]

Datasets: Every participating company provides granular data on their performed trips and type of transport (in case of multiple options such as bike, scooter, kick-scooter, etc). They expect to receive in return a monthly evolution of total number of trips by origin of generation, segmented by mode of transport, for all the areas of the city, to better understand where to place marketing actions or their fleets.

Limitations: The trips performed by each company should not be inferable from the generated dataset. There cannot be a data user that is not, in turn, a data contributor. The transformed dataset generated in the data sharing process cannot be used by third parties under any circumstances.

Ontology map: The data sharing agreement is composed of a single process, from data contributors (companies) to data intermediary (government) and back to the same data users (companies). There is no transformation certifier involved as there are no legal obligations but there is a process enforcer nominated by the parties for external audits on data processing by the data intermediary to ensure no leakage is present and original data is properly erased.

Process Companies which have technological capabilities perform first data erasing on the details of the trips that are not relevant (destination, user, etc) and send the data to the data intermediary. The data intermediary performs pooling by merging and aggregating the data by location and time. The data is then given back to the same companies and the original data provided to the data intermediary is erased.

Law of application Business secret protection applies to the entire set-up, as no external actors are allowed to see the data. No personal data is involved as user IDs are never shared.

Identification risks Personal identification probabilities are very low, so are costs, and overall risk is very low. Business identification risks are high, because while identification probabilities are moderate (if combining the data with the individual data for a given company with enough market share), identification costs are very high as they expose market detailed data of the companies.

Data transformations Each data contributor is in charge of erasing non relevant items from each data entry (destination, user data, etc) and data intermediary is in charge of aggregating data by origin location and time and mode of transport, erasing company ID numbers, and finally erasing the entire original provided datasets.

[76] See Martens and Zhao (2022) cited above

Agreement contract Each process of data sharing corresponds to a year, so companies can decline to share data in certain years. All the processes generate atomic contracts that are captured on a single agreement.

B2G2B: Mobility as a service (MaaS) [hypothetical]

Use case: All companies that offer mobility at different modes of transport need a platform that can generate a one-stop shop for the end user, to avoid competing for their attention and focus on offering an integrated solution for the better planning of urban trips. However, while all parties agree that a unique platform is required, everyone wants to own such a platform, and no interoperability is made possible due to that. Mobility companies have trouble accessing a large user base, and specific app developers have trouble accessing a large, truly multi-modal offer for their users. A public data intermediary can step in to solve such a dilemma, acting as a truly *neutral* actor and optionally benefiting from the generated data (see licensing case).

Datasets: On the one hand, every company provides full details live of the status of their offer, fleet position, price, etc to the data intermediary. The datasets are updated live on a continuous basis. On the other hand, companies that develop specialised UX and consumption apps of mobility for users provide requests of rides and mobility planning to the data intermediary, including trip details. The data intermediary matches requests to offerings and channels them to the user. If the user chooses one, the data intermediary puts in contact the lead generator and mobility operator for payment and execution.

Limitations: The only actor with full vision of both satisfied and unsatisfied transactions is the data intermediary. The other actors can only observe the demands channelled by them or to them. Optionally, the data intermediary may generate aggregated reports on market status to the data contributors (see previous case). No biases should be enforced by lead generators to promote certain operators.

Ontology map: In this case we have two types of data contributors (lead generators via apps facing the user and mobility operators), a data intermediary and the same data contributors that act as data users. Some or all of these will likely need to resort to technical data handler(s) due to the involved level of integration required. Because there are payments included in the process, both a process enforcer and transformation certifier will need to be appointed. The data sharing agreement is composed of two types or processes.

Process one In this case, both lead generators and mobility operators provide a continuous stream of data to the data intermediary and act as data contributors. They are then matched, and the offering details are given to the lead generator (data user) for the user to choose.

Process two: If a ride is chosen and a deal is closed, both the mobility operator and the lead generator notify it independently to the data intermediary, which matches both results.

Law of application The scenario is hypothetical and would deserve a detailed legal analysis, but it is to be expected that many laws will apply, from business secret and personal data protection to specific transportation ones.

Identification risks In this case, business secret risks in general are low, because only the involved parties in chosen trips observe the data to the trips where they are involved. The fleet status and pricing is never shared among operators, and only the relevant status and the minimal information for the user is shared to the lead generators. On the data shared, the probability of identification is very high (both personal and business details are known by the involved parties) but the cost is very low (as it is intended and accepted by both user and involved companies), so risk is moderate. There is however a significant risk of misuse via scraping to unlawfully obtain fleet information by posing as lead generator, but this can be addressed imposing usage fees, monitoring, and banning among other options.

Data transformations The only procedure required is data erasing of certain records that are required to know the basis for the lead generator from the mobility companies to provide complete offerings to the users.

Agreement contract Most likely a general agreement comprising both types of procedures should be generated, and then also each transaction should generate an atomic contract of data sharing. Being a hypothetical case, the details of the agreement follow from the other entries in the use case repository, so specific analysis should be carried out on those terms.

Annex 2: Technical annexes

Parametrised data transformations

Given one or multiple datasets, a data transformation is an operation that takes this input and outputs a new dataset, constructed entirely from the imputed data. In the context of urban data sharing, we only consider transformations whose intent is to decrease the granularity level of the input dataset (the rest of transformations possible are not interesting for this case, thus not susceptible to be coded into atomic contracts). We understand a dataset by a corpus of original data, even if it contains multiple tables, elements, columns, or fields.

Broadly, we identify two types of transformations, depending on the number of inputs. Individual transformations apply to operations performed on a single dataset, while collective ones apply to more than one. The latter case, for urban data sharing, only contemplates the pooling operation. That is, relating entries from different datasets (e.g. relating datasets of frequency of bus rides to a dataset of points of interest nearby the bus stops). Note that once a pooling from two datasets is performed, further individual transformations can be applied to it. In general, when combination of datasets is required, we suggest using a paradigm of transforming individually first the datasets, then pooling them, unless technical measures suggest otherwise (for instance, it is impossible to merge datasets that have previously been aggregated at identity levels).
Regarding individual transformations, we identify the following:

- **Aggregation:** operation of performing a (potentially normalised) sum over items along a given dimension. Dimensions can be temporal, spatial or categorical (among which the most relevant is that related to individual IDs).
- **Thresholding:** operating of computing whether a value lies between two boundaries (yes/no), it may be combined with an aggregation (count) to obtain a certain binning (e.g. how many bike trips we observe between 1 and 2 km?).
- **Sampling:** reducing the level of individual data by subtracting entries from it, either in a randomised way or by filtering them.
- **Randomising:** process of adding noise to the data, either by swapping certain entries, adding new fake ones, or altering certain entries by adding elements to them. Ideally, the randomisation effect should be controlled, so key properties of the data are maintained.
- **Mimicking:** Generating new entries that, while being fabricated, retain statistical properties of the original entries.
- **Erasing:** Not divulging certain parts of a given data entry (columns) that are not of interest for the considered use case.

The level at which one precisely describes the allowed data transformations depend on each data sharing atomic contract and the will of the actors sharing the data (and the constraints optionally imposed by the transformation certifier). In general, the more detailed the better, but at minimum they should include the dataset of origin, the type of transformation allowed and the parameters of each transformation (ranges, limits, etc).
All of the above procedures decrease the level of granularity of the input dataset, so makes identification of single points in them more difficult. The trade-off to pay,

obviously, is that the more aggregated the resulting information, the less useful it may be. Typically, when one wants to perform such operations, it aims at preserving certain statistical properties that contain the relevant information one wants to use. However, the usual problem is that one mostly does not know what those properties of interest are beforehand for each use case, so a complete exploration of the data is required. Note that, here, the relevance of a data intermediary that is in a position to do so, especially when pooling is required, is especially relevant.

Parametrised contracts

As justified in the technical section 4, we want to provide templates and modules that can (semi-)automatically create data sharing contracts that are readable by humans and machines. For this, we want to build on the work of the International Data Spaces Association (IDSA) and the Gaia-X initiatives, which have been working intensively on similar issues in the context of data sharing and data spaces over the last few years. The following is only a brief summary of the relevant work for our proposal, and a suggestion of how it could be further developed. A detailed description of the relevant work can be found in the International Data Spaces Reference Architecture Model[77] and the "Usage Control in the International Data Spaces"[78] position paper.

The approach allows data contributors to define access and usage conditions for the data they share. To make these conditions machine-readable, they are expressed using the Open Digital Rights Language (ODRL)[79], a W3C recommendation which specifies a vocabulary and data model for the description of digital and machine-readable contracts. Because not every data contributor is familiar with this language and its semantics and syntax, it is recommended to provide a user interface that allows the user to express and select predefined usage conditions in natural language and then convert them into the ODRL machine-readable format. A prototype interface developed by Fraunhofer for this purpose can be found here: https://odrl-pap.mydata-control.de/

The IDSA has identified certain recurring requirements for data access and data usage conditions of data contributors through the study and analysis of use cases in data sharing projects. For example, data contributors may want to define that the data they share can only be used in a certain time interval or restrict the use of the data to a certain location. These recurring patterns of data usage restrictions have been divided into 21 atomic templates called “IDSA policy classes”, which are explained in detail and expressed in machine-readable terms on GitHub[80].

For the technical implementation of a data intermediary, we propose to build on this preliminary work and take the 21 “IDSA policy classes” as a starting point. For the implementation of use cases, it should be assessed whether these policy class templates of data usage conditions cover the needs and requirements of the data

[77] https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/
[78] https://internationaldataspaces.org/download/21053/?tmstv=1692167531
[79] https://www.w3.org/TR/odrl/
[80] https://github.com/International-Data-Spaces-Association/IDS-G/blob/main/UsageControl/Contract/

contributors. To do this, the requirements of the data contributors must be mapped to the existing templates of the policy classes, selected, and the corresponding template fields, such as duration of use, location of use, etc., filled in (Annex 1: Use case repository examples). As mentioned, not only one policy class can be selected for each use case, but that any combination of policy classes is possible for each data sharing process. If requirements or needs for data access and usage conditions of data contributors arise, which are not yet covered by the IDSA policy classes, these can be easily extended. We anticipate the need for expansion on the one hand at the definition of the required data transformation processes. To address this, the grammar for possible data transformations presented in Annex 2: Parametrised data transformations should be used and translated into possible atomic terms. On the other hand, necessary extensions are conceivable for defining data sharing and usage conditions across multiple actors for B2G2X scenarios. To implement these conditions, the scenario should be broken down into individual conditions, which in turn are expressed as atomic terms, the sum of which ultimately describes the implementation of the entire scenario.
If one wants to implement this independently, it is necessary for the implementation to learn the ODRL language, and accordingly to express the new data usage conditions with ODRL to ensure interoperability with the other data usage conditions. An alternative, which we also recommend, is to engage in exchange with existing initiatives such as IDSA and Gaia-X and contribute the requirements to their open-source communities. Most likely, other initiatives and their use cases have similar requirements, which can then be implemented together with the community. Possible initiatives for information and knowledge exchange could be the Mobility Data Space[81], EONA-X[82], or the Data Spaces Support Centre[83].

Detailed architecture

Overview

The realisation and implementation of a technical data handler is very specific and strongly dependent on different factors, such as the existing know-how, already existing IT infrastructure, planned use cases, etc. In general, we recommend that all cities and regions begin by building on as much existing groundwork from related initiatives as possible. Moreover, the main rule for implementing data sharing scenarios should be to use the simplest solution available that meets the requirements of the case (i.e., the one that uses less components) In addition, as many common and open standards, and best practices as possible from the field of data management and data sharing should be used for the implementation. Examples of architecture design include describing interfaces using the RESTful paradigm and the OpenAPI specification[84], as well as communicating via HTTPS. Furthermore, it is useful to reuse existing vocabularies for describing (meta) data, such as DCAT[85], to ensure interoperability at the data layer.

[81] https://mobility-dataspace.eu/
[82] https://eona-x.eu/
[83] https://dssc.eu/
[84] https://www.openapis.org/
[85] https://www.w3.org/TR/vocab-dcat-2/

In the following we first describe the architecture modules which we see as compulsory. Depending on the requirements and envisioned functionalities of the use cases, there can be the need for the further optional modules, which we describe after.

Main modules

Identity manager

As described above, a key element in establishing trust in the data sharing process is distinct identity management. The identity manager takes care of the identification, authentication, and authorisation of users of the data intermediary. It ensures that organisations, individuals, machines, and other actors are provided with acknowledged identities, and that those identities can be authenticated and verified. This is essential for the secure operation of the intermediary and to prevent unauthorised access to data. Thus, each participating actor necessarily has an identity (which describes the respective participant on a technological level), is responsible for its fair use, and can use it for authentication. There are many ways to implement identity and access management, from centralised approaches such as OAuth 2.0[86] to decentralised approaches such as Self-sovereign Identities[87] technologies.

Agreement vault

To reduce the efforts necessary for the data sharing process, especially of contract negotiations, we propose to create an agreement vault. The structured management of agreements (and their machine-readable format) allows them to be stored, organised, and tracked in an accessible way to streamline future contract negotiation processes, improve cooperation between the actors involved, and thus reduce transaction costs in the future. The agreement vault contains historical information on all the agreements that have been put in place and is a reference for the currently enforced ones. It can thus be used for two functions: As a browsable reference to build new agreements, and as a registrar to validate allowed operations on data sharing processes.

Use case registrar

The use case register serves as a repository for organising, documenting, and managing use cases of data sharing use cases that each data intermediary has enabled. It can optionally be connected and federated with the network of data intermediaries for further volume. It serves two main functions. On the one hand, it helps other cities and municipalities browse similar requirements analysis, design, and implementation to their use cases. Each use case in the register should be described using the use case template card described in Annex 1: Use case template card. The use case register can also be extended to include other functions such as versioning and traceability if required. For the implementation of the first function of the register, at least in the beginning, a relational database is suitable.

The second function relates to keeping an original reference to uniquely identify data access claims. Same as with the agreement vault, any data access claim to the system

[86] https://oauth.net/2/

[87] https://www.w3.org/TR/did-core/

must incorporate a reference to existing agreement, use case and information vault (see below).

Activity logger

The activity logger is a technical component which will log and store every data transaction happening through the intermediary. By this the activity logger provides the means for tracing and tracking in the process of data provision, transformation, and use. At minimum the activity logger should log each time which data contributor provided which data for what purpose, linked to which agreement, term, and use case, what transformations and/or pooling were applied to the data, and who used these data for what use case. In addition, the logging of further details of the data, such as data quality is conceivable. The activity logger thereby acts as the basis for many important functions, from identification of the lineage of data to audit-proof logging of transactions. It has a capital role, as it can be used to ensure trust in the system, and eventually demonstrate misuse by actors (and its associated consequences). There are different ways to implement this activity logger, e.g. through a (simple) log file, a (potentially distributed) database, or cryptographic technologies.

Auxiliary modules

Data checker

The data checker is a software tool that checks the shared data for quality and other predefined rules, such as data formats and agreed standards (present in the sharing terms). This is necessary to ensure data quality as well as data interoperability. The implementation depends largely on the connected data sources, their data formats and quality, as well as the required target formats and quality. In general, however, one should try to automate as many validations and transformations as possible to ensure scalability. The data checker also implements procedures to validate that proper anonymisation techniques have been applied to the original datasets, if this was provisioned in the sharing agreements, to ensure separation of legal responsibilities (the data intermediary might not be required for individual anonymisation techniques on individual datasets). There are already some widely used data checkers as open-source solutions, which can be used for initial implementations or as guidance. See e.g. pydantic, pandera, great expectations.

Data transformer

The data transformer is responsible for the parametrised data transformation processes described above. Consequently, the data transformer will transform a shared dataset according to the agreed data sharing terms (ranges, limits, etc.) and transformation process (aggregation, randomisation, etc.). There are different approaches and methods for the implementation of such a data transformer which depend, among other things, on the required complexity of the transformations, scalability requirements, and the existing technology stack of the technical data handler. The transformer should be implemented in such a way that the entire transformation is broken down into individual, traceable steps. This allows the transformation to happen successively, so that each step can be logged and verified by both the technical data handler and the transformation certifier. One possible implementation is to write a custom script in familiar programming languages such as

Python, Java, or SQL. This implementation method provides flexibility in development and implementation with respect to data transformation logic requirements.

Info vault

The information vault is an optional module that may be used to store calculations and analyses that have already been performed in the past with shared data, as long as its associated sharing terms allow for it. This is especially necessary for computationally intensive data transformations and analyses, as it allows faster retrieval or analysis for other data recipients. The simplest way to store these results is in a relational database. However, for more advanced scalability or availability requirements, other storage technologies, such as non-relational databases, are conceivable.

Annex 3: Data sharing FAQs for cities: What elements are required when?

This final technical section is meant as a guide for administrations wishing to implement the proposed tools present in this document, and it is written in the form of a Q&A document.

When does one not need a technical data handler?
A technical data handler is not needed either in cases of simple, untransformed, one-to-one data sharing, or, when the data intermediary is an organisation mature enough to apply data transformations themselves. For smaller cities, we suggest this to be externalised to a non-governmental entity, while for large cities, we suggest the city take this role, commissioned by the data intermediary.

When are data transformations needed to be implemented by the technical data handler?
The categorised data transformation procedures are only needed in three specific cases. The first case applies when there is pooling involved in the data sharing, that is, when datasets need to be first combined and then aggregated, and that aggregation cannot be done by the data user itself, due to confidentiality concerns. In this case, the data intermediary is allowed by the data contributors to commission this activity to the technical data handler (or perform it itself if it has sufficient technical capabilities).
The second case arises when the data contributors do not have enough technological knowledge to implement original dataset protection measures that apply solely to their contributed datasets (anonymisation techniques) and accept delegating them to the technical data handler.
The last case applies when, due to simplicity or technological reasons, simple aggregation procedures are delegated to the technical data handler to avoid the data contributors to perform multiple, different aggregations to generate different versions of information over the same datasets for different atomic contracts.
We suggest by default, unless strict confidentiality aspects demand it, to delegate data transformations to the data intermediary (and if needed, the technical data handler) to simplify set-up and negotiation among the parties.
Of course, the modules of data pooling and data transformation are not needed in the cases described in this paragraph.

When do I need an information vault?
Whenever the data technical handler must generate different versions of the data for different actors and/or must combine data from different sources, an information vault is required. This can be an actual data storage (if the process of generating the versions of the data is lengthy) or just a script storage containing the instructions to do so on the fly.

When is a transformation certifier required?
The presence or absence of a transformation certifier does not affect the architecture, as certification is meant as a one-shot event per transformation proposed. Having said that, any data sharing scenario involving data that might be related to individuals would

certainly require their involvement (and almost all urban data sharing scenarios are like that).

When is a process enforcer required?

The choice to delegate authority to a process enforcer is up to the actors involved in the data sharing and part of the negotiation process, it is by no means a required actor. The implication of a process enforcer from a technological perspective is basically the need to provide all powerful read access to all the involved infrastructure in the data sharing to a privileged actor. Such an actor must ensure the correct custody of such powerful credential access rights.

What technologies should I choose?

As stated earlier, on the implementation side a strong preference for open source should be taken. However, not only open-source implementations should be enforced, but also mechanisms to ensure their usability and iteration: Open repositories, modern development pipelines, documentation, and excellent code engineering practices. It is the only way to ensure that small cities can profit from cumulative investments by the larger administrations.

Moreover, care should be taken to balance technological novelty with practicality and use. This means that experimental or beta technologies, packages, and programming paradigms should be avoided when established options provide enough functionality. In public administration scenarios, management of legacy systems can be especially hard, and so choosing too complex solutions can lead to difficulties in maintenance and HHRR. This is more important for large cities than small ones, which are susceptible to externalised data sharing and as such, may have less inertia to implement changes or change providers.

Finally, regarding the use of standards, care should be taken when selecting among competing options. Standards emerge from a combination of political, community and market consensus, and so options that present implicit or explicit technological lock-in should be avoided.

How would such a set-up work in practice?

For an urban data sharing scenario, typically administrations would rely on a data intermediary that would delegate the data processing to an external technical data handler (or do it themselves), that would provide the data transformed at the agreed level of detail to the data users, one of which may be the city administration, but others are possible. Then, one must differentiate between large and small cities.

For large cities, to promote technological sovereignty, the ideal situation would be that the department in charge of data governance (CDO) could act as data user, and then use and govern this data for internal and/or external purposes (open data). This might or might not involve hiring third parties or using internal resources for building services that use this data. If the usage is meant for new use cases not covered in the existing data sharing agreement, new terms must be added to it.

For smaller cities, this might be more difficult, so either the entire role of data governance may be externalised, or it can be done by the technical data handler. In the latter case, the city facilitates contacts directly with its providers of specific solutions that might depend on such data (open-data portal, business intelligence, specific city apps).

In all cases though, the control and governance of the data should entirely reside within specialised knowledge of the administrations, be them large or small, due to their high (and increasingly important) relevance in public affairs. And any involved third party will not be free nor possess any rights over the data processed, besides those authorised by the scope of work mandated by the city to do.

Finally, notably, the technical data handler role may be taken by the city administration itself (if all parties agree and enough technological capabilities are present), although this might be a case for larger cities. Another option is for cities to pool together and fund a powerful public, not-for-profit entity that may take this role altogether, bundled (or not) with that of the data intermediary, although this lies outside the scope of this blueprint.

With which use case can/should I start?

We suggest started mapping the most relevant questions and challenges to build potential use cases from them. Once this is done, we advise to always choose the simplest use case that can provide value. This means choosing a case among the candidates that is complicated enough to test the set-up of data sharing terms, while at the same time is simple enough that it can be achieved in reasonable time involving the minimum number possible of actors. Even if the chosen case is not the one addressing the most pressing issue, it will be a good choice as an innovation tool. The most pressing need is always the most impactful use case, but it is also usually the most complex and one that carries most expectations, and thus should never be applied first without experience.

Annex 4: What we learned about using the challenge format for our experiment

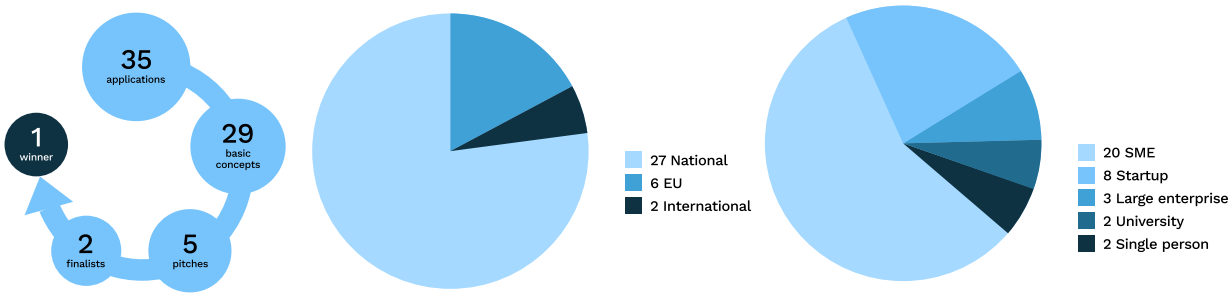
For a final assessment of the format of an urban data challenge, different aspects and targets are to be considered:

1) **Local experiments such as the Urban Data Challenge serve as a basis for gaining practical insights on B2G2S data sharing in city contexts.**
   As often in complex and innovation projects, gaining insights is a key aspect of the undertaking. Although ideally, all initial hypotheses and plans are proven to be valid and correct, that is not how it usually works. However, also the aspects that did not work out as planned are very valuable for the innovation and learning process. Experiments, carried out with meaningful use cases and with the ownership of city officials and high-level policy commitment, are essential when looking to change policies or to regulate in the field of digital and technology.

2) **The urban challenge as a format can successfully solve real challenges of city administrations.:**
   Although the collaboration of the University of Bremen, the winner of our competition of ideas, and the Authority for Transport and Mobility Transition (BVM) is ongoing and final results are outstanding, an initial evaluation shows, that the winning team adds new capacities and competencies to the process of analysing the effect of a pop-up bike lane specifically and micromobility flows in Hamburg in general. A key factor for the project has always been to ensure an actual use case for the experimentation to learn in a real environment and demonstrate the potential of using data for the public interest.

3) **The urban challenge format very successfully stimulated and involved the innovation ecosystem.**
   We received 35 applications from diverse backgrounds (Figure A4.1) - academia, start-ups, large companies, consultancies – and from Germany, Europe, and the USA. Subsequently, 29 of them successfully submitted a fully eligible set of documents including their basic concepts.
   These numbers are, also compared with similar initiatives from for example Barcelona, very high.

Figure A4.1: Overview of Urban Data Challenge participants



4) The urban challenge format was fruitful in the case of the Urban Data Challenge, but carries high bureaucratic burdens and therefore is only partially practical for experimenting

We chose the format of an urban challenge with the ambition of organising a lean and appealing competition of ideas with low barriers to participate. However, the initial expectation and understanding that a complex procurement process could be bypassed, as our challenge prize money (40,000 EUR) was below the threshold ("Unterschwellenbereich"), proved to be wrong. Instead, we learned that an official procurement process was still required, as the fact that the prize money is below the threshold does not mean that procurement processes can be neglected, but mainly changes the options of legal procedures for the participants in the process. To structure and guide us through the procurement process, we thus decided to collaborate closely with Dataport, an IT service provider for the public administration in Hamburg. Nevertheless, the process turned out to be very time-consuming for all stakeholders involved (e.g. the different departments from the City of Hamburg, The New Institute team, lawyers). Most importantly, the process for the challenge participants was not as straightforward and easily accessible as anticipated in the beginning. Our recommendation for the administration would therefore be to make greater use of the leeway that procurement law provides to overcome the somewhat conflicting goals of procurement processes and the promotion of innovation.

Annex 5: Members of the Data Commons Working Group

- **Francesca Bria (Chair):** Director The New Hanse
- **Renata Avila:** Open Knowledge Foundation
- **Aline Blankertz:** Data economist Wikimedia
- **Malcolm Bain:** ID Law Partners
- **Marco Ciurcina:** NEXA Center for Internet and Society
- **Fernando Fernández-Monge:** Bloomberg Harvard City Leadership Initiative
- **Adrian Fiedler:** Free and Hanseatic City of Hamburg
- **Maximilian von Grafenstein:** Einstein Center Digital Future
- **Moritz Hennemann:** University of Freiburg
- **Rainer Kattel:** UCL London
- **Paul Keller:** Open Future
- **Raffaele Laudani:** City of Bologna
- **Henriette Litta:** Open Knowledge Foundation Germany
- **Marina Micheli:** JRC European Commission
- **Geoff Mulgan:** UCL London
- **Paul Nemitz:** European Commission
- **Boris Otto:** Fraunhofer ISST
- **Dominik Piétron:** Humboldt University Berlin
- **Oleguer Sagarra Pascual:** Dribia Data Research Barcelona
- **Maria Savona:** Luiss University Rome
- **Linnet Taylor:** Tilburg Institute for Law, Technology and Society
- **Stefaan Verhulst:** The GovLab New York

Experts invited to contribute to ad hoc sessions:

- **Pau Balcells:** City of Barcelona
- **Joshua Gelhaar:** Fraunhofer ISST
- **Justin Nogarede:** Friedrich-Ebert-Stiftung
- **Stefania Paolazzi:** City of Bologna
- **Kay Pöhler:** KfW Bankengruppe
- **Jan Pörksen:** Free and Hanseatic City of Hamburg
- **Leevi Saari:** University of Amsterdam
- **Sille Sepp:** MyData Global

Acknowledgements

The authors would like to thank many organisations and people for their support:

The members of the Data Commons Working Group and all experts invited to contribute to ad hoc sessions.

The New Institute and its networks:
Erck Rickmers, Sebastian Hofer, Wiebke Hallerberg, Adriana Groh, Katharina Meyer, Ana Paola Lopez, Georg Diez, Christoph Gottschalk.

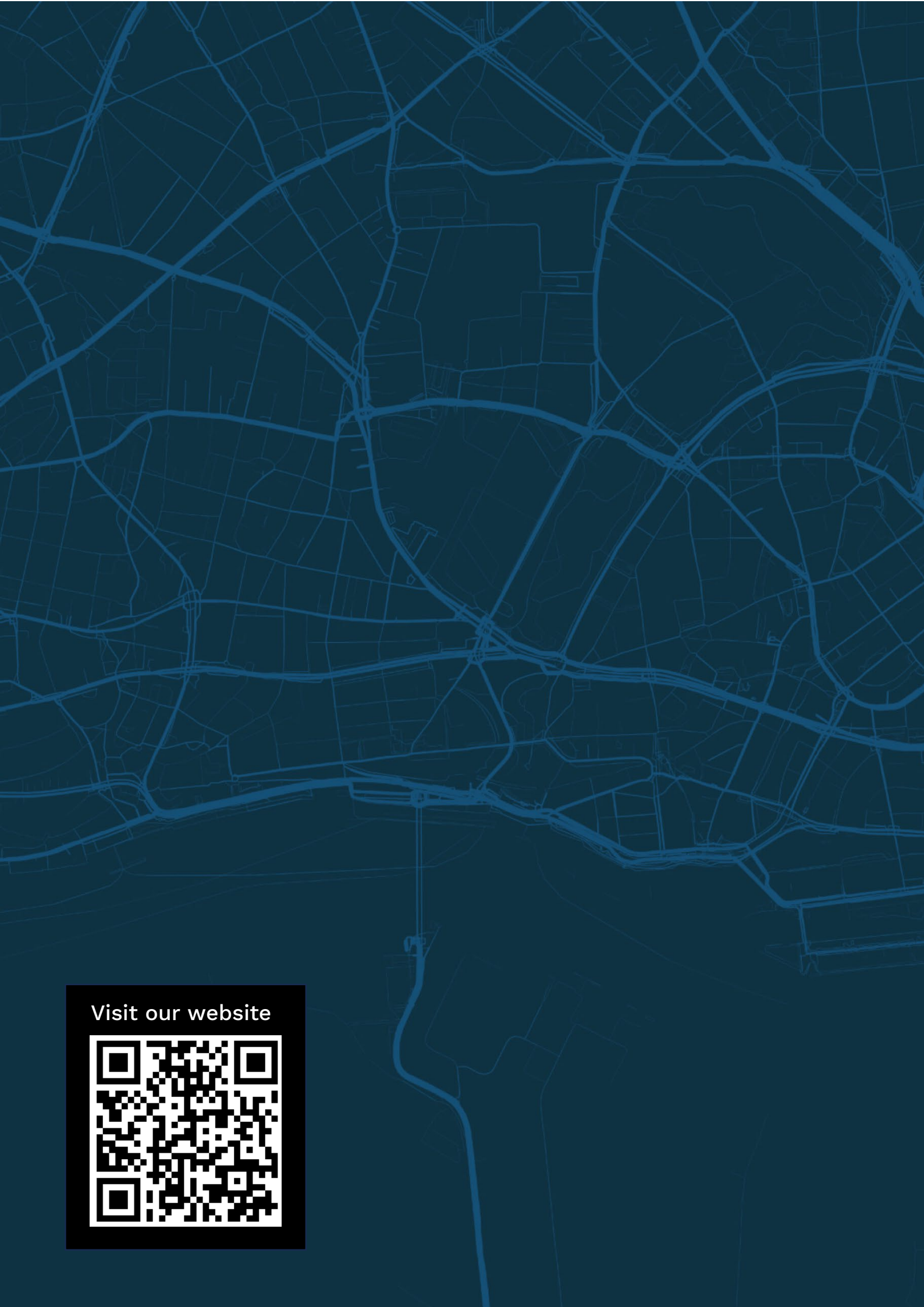
Our partners of the Free and Hanseatic City of Hamburg:
Jan Pörksen, Christian Pfromm, Matthias Wieckmann, Adrian Fiedler, Christian Alwardt, Elisabeth Kaufmann, Diether Schönfelder, Malte Noga, Mareike Behrendt, Andrea Weidinger, Pierre Gras, Michael Fischer, and the Hamburgische Beauftragte für Datenschutz und Informationsfreiheit.

Our partner companies/ data contributors:
IOT Venture and Mario Schmitz
Bolt and Natascha Spörle

Published by The New Institute.

How to cite this report: Bria, F. et al., *Blueprint Governing Urban Data for the Public Interest,* The New Institute, Hamburg, 2023, doi:10.17605/OSF.IO/FT4ZX

Layout: CDLX GmbH and The New Institute



Visit our website