Making Sense of the Big Data Mess: Why Interdisciplinarity Matters in Smart Cities

Frechen, Niklas; Heger, Pauline; Bieber, Christoph; Hendawy, Mennatullah

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Frechen, N., Heger, P., Bieber, C., & Hendawy, M. (2023). Making Sense of the Big Data Mess: Why Interdisciplinarity Matters in Smart Cities. *easy_social_sciences*, 69, 45-52. https://doi.org/10.15464/easy.2023.12

easy_social_sciences 69 ▪ 2023

doi: 10.15464/easy.2023.12

Making Sense of the Big Data Mess

Why Interdisciplinarity Matters in Smart Cities

Niklas Frechen, Pauline Heger, Christoph Bieber & Mennatullah Hendawy

Smart cities use vast amounts of (big) data, often creating what we call an urban "data mess". In this article, we show the **diversity and complexity of data** *that make up this mess and outline examples of urban data processing. Furthermore, we point out problems with the sector-specific perspective that is usually taken when dealing with smart cities. We argue that a collective way of dealing with data across sectors and disciplines needs to be found. To achieve that, we advocate for* **more interdisciplinary cooperation** *between different disciplines and stakeholder groups. The Pandemic Recovery Dashboard of the City of Los Angeles gives a first impression of how this could work. We aim to show that approaching data in smart cities from an interdisciplinary angle may help deal with the data mess in smart cities – both for researchers and city developers.*

„Smart Cities" stützen sich auf große Datenmengen („Big Data") – wobei die unterschiedlichen Daten häufig in ungeordneter Form vorliegen (engl.: „data mess"). Im Beitrag widmen wir uns dieser **Diversität im städtischen Datenbestand** *und skizzieren Beispiele urbaner Datenverarbeitung. Dabei verweisen wir auf Probleme und Herausforderungen einer engen, an einzelne Bereiche gebundenen Datennutzung: Aus unserer Sicht fehlt bislang ein gemeinschaftlicher, sektorübergreifender Ansatz zum Umgang mit Smart-City-Daten. Aus diesem Grund sind* **mehr interdisziplinäre Kooperationen** *erforderlich, d.h. die Zusammenarbeit unterschiedlicher Disziplinen und Stakeholder-Gruppen. Das Pandemic Recovery Dashboard der Stadt Los Angeles gibt einen ersten Eindruck davon, wie urbane Daten erfolgreich genutzt werden können. Wir argumentieren dafür, dass Daten in Smart Cities am besten in ganzheitlicher Perspektive bearbeitet und der städtische Datendschungel so übersichtlicher gestaltet werden kann – für Wissenschaft und Praxis.*

Keywords: smart city, interdisciplinarity, big data, urban data, civic tech

Cities have been of scholarly interest for a long time and lots of metaphors have been used to describe the urban scenery: cities have been conceived as organisms, as nature, as machines, as theatres, or as a form of memory. More recently, the increasing use of technology and the collection of data has brought about ideas of the city as a computer, an intelligent machine, or even a cyborg. In urban planning and city politics, the term "smart city" has become increasingly popular to describe this view.

The multi-faceted landscape of smart city projects all over the world is an interesting field not only for city planners but also for the growing number of people living in cities as well as for scholars from various disciplines being confronted with new objects for research. This article deals with a key element of modern city life: the collection, storage, and processing of many kinds of data to improve cities in various regards. The basic idea focuses on improving knowledge about what determines the life of city dwellers. As modern cities have grown into complex structures, many different sources provide data that can be tracked, stored, and processed – but in the first place, they create a complex and confusing mass of unstructured material. This situation holds lots of challenges for city administrations and politics as well as for researchers. Modernization of long-standing organizational structures is needed as well as expanding interdisciplinary efforts in data and city management.

What is a Smart City?

Software studies scholar Rob Kitchin defines a smart city as "one that can be monitored, managed, and regulated in real-time using ICT infrastructure and ubiquitous computing" (Kitchin, 2014, p. 132). Put simply, cities often qualify for being called "smart" when they use digital technology, such as security cameras, pollution sensors, traffic meters, or other specialized devices, to collect different sorts of data (see the following section). These data are used further, especially for processes of urban policymaking, e.g., for city officials' decisions about what services in terms of health, transportation and security will be offered to citizens. In most cases, plans for making a city smart (called *smart city strategies*) form the next step in a long history of city modernization. The overarching goal of many of these strategies is to use data-based technologies to improve the quality of life for their residents (Al Nuaimi et al., 2015).

When we look at the smart city strategies published by different city governments[1], we

»Cities are often called "smart" when they use digital technology to collect different sorts of data.«

notice that there are often specific visions the smart city tries to provide, indicating a certain "sense" as per Kitchin's (2014) words. Sometimes, these visions are called *smart city narratives* – which can be understood as a "story" a city wants to tell during the process of digital urban modernization. In Germany, for example, depending on the respective emphasis of the smart city strategy in place, we notice that promoted visions include *traffic-smart* cities (cf. Hamburg), *energy-* or *climate-smart* cities (cf. Paderborn), and sometimes *administration-smart* cities (cf. Nuremberg). As shown in Figure 1, certain narratives are closely related to certain applications and hence require specific kinds of data (e.g., traffic and transportation data would be very relevant for traffic-smart cities), as we will show in this article. Overall, while narratives, goals, progress, and implementation already differ greatly between smart city projects, there is even more variety when it comes to the kinds of data that are used and needed in these projects. In the next section, we elaborate on these different kinds of data.

What Kind of (Big) Data do Smart Cities Use or Produce?

Data is an essential part of smart city projects (e.g., Kitchin & Dodge, 2011). In a general sense, data refers to (collections of) many different pieces of information about something. However, there can be different ways in which this information is represented in terms of meaning, format (e.g., in numeric form, text-

1 Usually, cities accompany their measures with extensive documentation and strategy papers, outlining the basic goals of smart city activities. These are often publicly available on the cities' websites (e.g., for Los Angeles, Barcelona, and Bochum, to name a few).

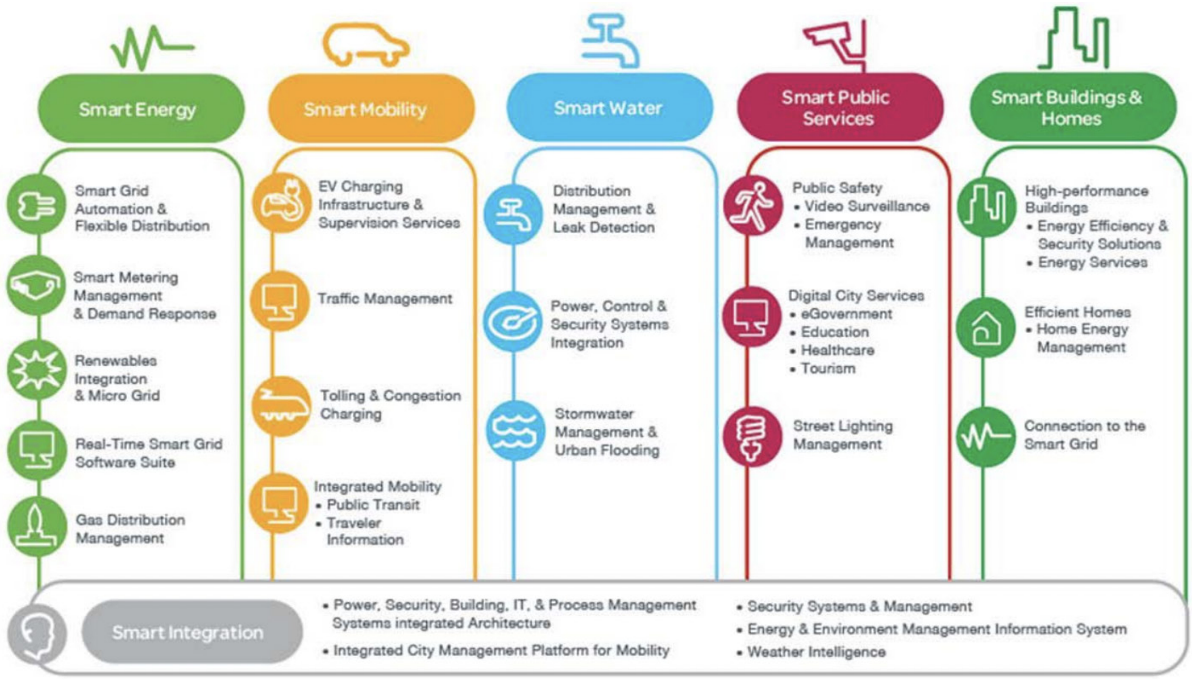

Figure 1 Smart City Services and Solutions. Image Source: Aoun, 2013, p. 9.

based, or otherwise), what units of measurement are used, and how the data is stored and processed. Additionally, the sources of the data themselves can vary greatly. As the data discussed here is produced by all technology implemented in cities, the amount and variety of data are extremely high. In other words: smart city data is big data.

Big data is often distinguished from regular data by its characteristics of *volume*, *velocity*, and *variety*, commonly referred to as the "3 Vs". These categories were introduced by Laney in 2001 as a framework for defining what constitutes big data. Volume refers to the massive amounts of data generated, velocity to the speed at which it is accumulated, and variety to the diversity of the data collected (Laney, 2001). While some of the big data is produced by humans both directly (e.g., election results, participation data) and indirectly (e.g., platform usage data collected by providers), there is also "data produced by objects", which derives from sensors, cameras, RFID tags, or Wi-Fi signals (Flyverbom & Madsen, 2015, pp. 124, 131). In a "mobility-smart" city, for example, this can include data from pollution sensors on busy roads, or the passenger data from public transport. By comparison, in an "energy-smart" city, this can include data from smart grids, or feed-in data from private solar power systems. Data produced by objects is an important driver of data collection and processing by administrative or other actors and also fuels the data resources of smart city projects.

Within the masses of data that a smart city creates (and needs), one can distinguish between various *data sectors* that smart cities attempt to develop (e.g., smart mobility, smart administration, smart energy, etc.). Until today, we have observed that data in smart cities has usually been dealt with from a sector-specific perspective, both by researchers and smart city planners. These sectors, in turn, involve many different *data types,* such as environmental data, geographic data, traffic data, data from politics and administration, and statistical data. An illustration of this sector-based approach and the associated data types is shown in Figure 2. The division into data sectors is usually also reflected in the structure of city governments and institutional divi-

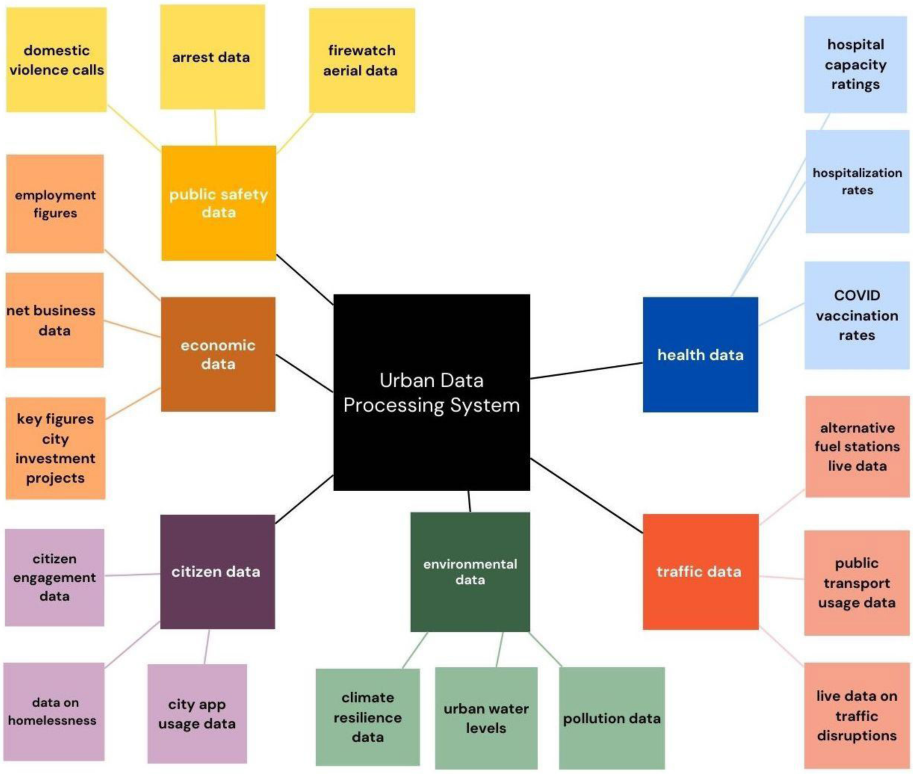

Figure 2 Exemplary presentation of different data sectors and corresponding data types produced in smart cities (developed by the authors).

sions (cf. Bundesministerium des Innern und für Heimat, 2012). Thus, data sectors are often directly connected to specific policy domains. The scope and priority of these domains (and of the associated data sectors) depend on the chosen narrative of the city.

Even though generating data in smart cities may be common practice by now, using it can be challenging. To be ready for interpretation by experts, researchers, or policymakers, these masses of data must be processed—which means *integrating* and *aggregating* the data and combining it with already existing data (Bischof et al., 2014). Integrating data means bringing together data from different sources, such as different sensors or cameras, into a single system or platform. Aggregating data, on the other hand, means summarizing or combining data into a more manageable form. This can, for example, mean grouping data based on specific criteria, such as time periods, geographic regions, or demographics.

However, processing the different sorts of data for city management and decision-making is only the first step in a complex challenge. In the next section, we show that the variety of sectors and types of big data involved creates many problems and thus results in what we call an urban “data mess” in smart cities – which cannot be dealt with from one single perspective.

What is the Problem With Disciplinarity When Looking at Smart City Data?

A symptom of the data mess is that big city data is often not used effectively (Hashem et al., 2016). This can lead to insufficient policy making, like a city recording traffic data, but failing to use it to improve traffic light circuits. While investigating such failures, researchers find that tackling the 3 Vs is often unsuccessful. Smart city data comes in huge volumes and is also varied in terms of type (e.g., traffic, weather, noise, and pollution data), source (e.g., different types of sensors, meters, etc.), and quality. Furthermore, such data is produced in different velocities, which range on a scale from static (only ever updated manually) to dynamic (continuously and automatically updated), can be provided in differing measurement units or formats, and may differ in structure (Osman, 2019; Bischof et al., 2014).

In addition to these challenges, the often fragmented structure of city governments with its many boards, departments, commissions, and bureaus leads to a large number of data silos that have to be re-connected within the structure of a citywide data catalog.[2] Examples of this problem can be seen in city-based data repositories, often called *urban data platforms,* where streams from different data sectors are collected and oftentimes only loosely structured by several non-standardized topics, categories, or formats[3]. Due to the diverse set of stakeholders involved in city politics, data-related expertise often remains scattered. There often is no exchange across *disciplines* (i.e., areas of expertise, such as sociology, politics, environmental studies, computer science, biology, etc.) due to a certain historical preference for *disciplinarity,* meaning a focus on a specific order of knowledge (discipline), where people share a common language, specific theories, and methods and, in turn, a focused and thus limited scope.

Within the context of big data and smart cities, this means that somebody trained in data science, for example, could have the expertise to handle environmental data from a technical point of view but would need the domain expertise of meteorologists, environmental engineers, and others to interpret these data. In turn, for acting upon the insights gained, there would be a need to interact with politicians and/or citizens to put the data to beneficial use. Also, some interactions between different issues (e.g., environmental problems arising from economic issues) might not be visible if only looked at from one specific disciplinary viewpoint. Thus, cooperation between several different disciplines is necessary to cover all the experience, interests, and skills needed to gain a complete overview (and to know what to do with it). Otherwise, urban innovations, interventions, and scientific studies will be limited to a small fraction of the technological, social, and political issues that arise from current narratives of smart city development.

How Can Interdisciplinarity Help to Deal With the Data Mess?

Interdisciplinarity, understood as the integration of methods and knowledge from various disciplines by combining different approaches (Stember, 1991), is an important lead for understanding the diversity of data and the complexity of smart cities. By bringing in diverse experts, more complex situations and challenges, such as the management of urban data, can be tackled (Lemos & Morehouse, 2005).

To assemble an adequately interdisciplinary team that can make sense of a city's data mess, a potential approach is to map the different data types that a smart city needs and/or produces and start from there to identify the expertise needed to effectively develop the project. In this regard, an overview as presented in Figure 2 can be a starting point that shows that some of the necessary experts can come from various disciplines, such as public safety, public health, data science, traffic management, environmental studies, social sciences, etc. An additional asset in covering diverse backgrounds and views for navigating smart city projects can be interdisciplinary persons whose backgrounds already lie at the intersection of disciplines, such as computational planning, feminist geography, and

2 The Los Angeles strategy, for example, envisions the central data repository as the "LA Data Lake" (Ross, 2020, p. 24).

3 These catalogs can provide an (incomplete) overview. Many cities or regional networks develop their own versions of open data access points (two examples: City of Los Angeles: https://data.lacity.org, Ruhr region: https://opendata.ruhr).

> ***Cooperation between several different disciplines is necessary to cover all the experience, interests, and skills needed to gain a complete overview.***

philosophy of technology, for example.

The most important argument for getting people with different backgrounds to work together is that smart cities can only be completely understood when seen from a *socio-technical* perspective. Smart cities cannot be seen from a purely technical standpoint, but also need to be looked at from a social one – cities are social spaces that are inhabited by (inherently social) human beings, after all. The technical view might focus on the practical implementation of new technologies, such as sensors or transportation schemes. Meanwhile, the social view includes the needs and challenges of the city's inhabitants, potential (non-technical) solutions, and the effects that new technologies have on citizens. To see both sides, smart city developers and researchers need to assemble teams that include experts from both social and technical sciences, e.g., from sociology, philosophy, psychology, political science, urban planning, engineering, economics, computer science, and others. For an interdisciplinary project, it is crucial to create a space for dialogue and collaboration among experts from different sides. That way, smart city development can, for example, produce technical solutions for social problems (cf. Trencher, 2019).

Finally, smart city research and development should include the interaction with city decision-makers (i.e., city officials; for implementing changes in practice) as well as the perspectives of the general public (i.e., citizens; for feedback and information). This very broad concept of interdisciplinarity includes different skills and expertise in a cooperative and socio-technical view by assembling diverse teams in line with the different kinds of data as well as bringing in political decision-makers and involving the general public.

One illustrative example for the implementation of interdisciplinary cooperation in smart city contexts is a recent project from the city of Los Angeles that was created during the COVID-19 pandemic, which we will present in the next section.

Example: Making Sense of Pandemic Data in Los Angeles

The case of Los Angeles is instructive, as the city made use of a broad array of accessible data to inform decision-makers while developing strategies to overcome the challenges of the COVID-19 pandemic and the resulting shutdown of large parts of the city. By pooling data from various sectors, the *Los Angeles Data Team* created a *Pandemic Recovery Dashboard* with datasets from four categories regarded central for a swift recovery from the crisis: COVID, crime, economy, and homelessness[4]. Figure 3 displays the extraction of pandemic-related datasets from the LA data catalog.

Los Angeles' *Pandemic Recovery Dashboard* shows a multi-domain approach seeking to integrate a diverse set of data and the accompanying expertise. To combine the information, domain-specific expertise had to be connected across several data sectors. The backbone was the cooperation of the Los Angeles Department of City Planning and the Information Technology Agency. Setting up a repository for city data (*Open Data Portal*) and a public platform for geospatial data (*GeoHub*) improved cooperation across administrative sections. Directly working under the mayor[5], the *Los Angeles Innovation Team* developed new ideas, infrastructures, and processes for civic design. Ideation sessions hosted by the team brought together civil servants from different departments, elected officials, and external partners like designers, urban planners, and

4 The dashboard data is published at https://data.lacity.org/stories/s/afkw-g9zz (accessed January 23, 2023).

5 The *Los Angeles Innovation Team* served under mayor Eric Garcetti (2013-2022) as an example for modernizing urban politics. The unit was not continued under the new mayor Karen Bass.

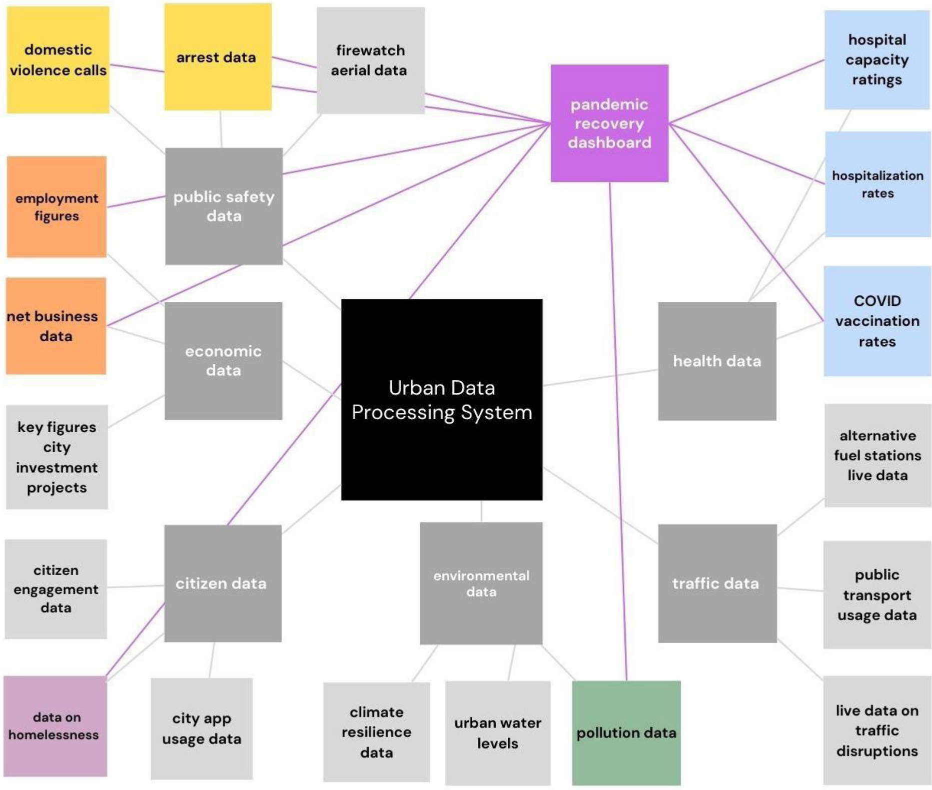

Figure 3 Sources for the Pandemic Recovery Dashboard within the Los Angeles Data Catalog (developed by the authors). Extracted data is shown in colors.

software developers. Among other projects, the team was responsible for the Daily *Los Angeles COVID-19 Data Summary* by connecting various sources from the city government with material from departments of the surrounding Los Angeles county structures. By using forms of online storytelling, Los Angeles' digital COVID-19 response connected the administrative efforts to the broader public. Publishing data-driven articles about crime trends in the city (Zhong et al. 2022) or a "story map" on the efficiency of the rental assistance program (Alcazar & Zavala, 2021) emphasized the possibilities of interdisciplinary cooperation. Nevertheless, projects like the *Pandemic Recovery Dashboard* remain somewhat incomplete by not making use of all the datasets available due to limited resources or the experimental status of new, inter-divisional teams.

Cooperation Beats Data Mess

In this article, we have argued that interdisciplinary cooperation should be a main approach both when studying smart cities and when developing them in practice – and that it could be helpful to shift the public debate about smart cities towards a more socio-technical perspective. Bringing in the expertise of diverse disciplines to the creation and study of smart cities allows for a better understanding and utilization of the data gathered in urban environments. Future research should explore how interdisciplinary cooperation can be effectively achieved to deal with the smart cities' data mess and how interdisciplinary cooperation in smart cities can be related to policy and practical implications. As of now, we conclude that interdisciplinary cooperation has the potential to benefit everyone involved in the process of handling the "big data mess" of smart cities – be it researchers, city planners, or city residents.

References

Alcazar, I., & Zavala, S. (2021, August 16). *ERAP evaluation. Evaluation phase 1 of the Emergency Rental Assistance Program*. Los Angeles Department of Housing. Retrieved August 31, 2023, from https://storymaps.arcgis.com/stories/0c59272161634bc69f3e0dd077fcf41b

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications, 6*(1). https://doi.org/10.1186/s13174-015-0041-5

Aoun, C. (2013). *The smart city cornerstone: Urban efficiency. Schneider Electric White Paper.*

Bischof, S., Karapantelakis, A., Nechifor, C.-S., Sheth, A. P., Mileo, A., & Barnaghi, P. (2014). *Semantic modeling of smart city data.*

Bundesministerium des Innern und für Heimat (2012). *Open government data Deutschland*. BMI.

Flyverbom, M., & Madsen, A. K. (2015). Sorting data out: Unpacking big data value chains and algorithmic knowledge production. In F. Süssenguth (Ed.), *Die Gesellschaft der Daten: Über die digitale Transformation der sozialen Ordnung* (pp. 123–144). Transcript.

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E., & Chiroma, H. (2016). The role of big data in smart city. *International Journal of Information Management, 36*(5), 748–758. https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Kitchin, R. (2014). Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy, and Society, 8*(1), 131–136. https://doi.org/10.1093/cjres/rsu027

Kitchin, R., & Dodge, M. (2011). *Code/Space: Software and everyday life*. MIT Press.

Laney, D. (2001). *3D data management: Controlling data volume, velocity, and variety*. Meta Group.

Lemos, M. C., & Morehouse, B. J. (2005). The co-production of science and policy in integrated climate assessments. *Global Environmental Change, 15*(1), 57–68. https://doi.org/10.1016/j.gloenvcha.2004.09.004

Osman, A. M. S. (2019). A novel big data analytics framework for smart cities. *Future Generation Computer Systems, 91*, 620–633. https://doi.org/10.1016/j.future.2018.06.046

Ross, T. (2020). *SmartLA 2028. Technology for a better Los Angeles*. Los Angeles. https://ita.lacity.org/sites/g/files/wph1626/files/2021-05/SmartLA2028%20-%20Smart%20City%20Strategy.pdf

Stember, M. (1991). Advancing the social sciences through the interdisciplinary enterprise. *The Social Science Journal, 28*(1), 1–14. https://doi.org/10.1016/0362-3319(91)90040-B

Trencher, G. (2019). Towards the smart city 2.0: Empirical evidence of using smartness as a tool for tackling social challenges. *Technological Forecasting and Social Change, 142*, 117–128. https://doi.org/10.1016/j.techfore.2018.07.033

Zhong, J., Gui H., Kotamreddy, H., Lew, A., Sherzai, A. (2022, October 4): A data-driven exploration of crime trends in Los Angeles. *DataLA*. https://medium.com/datala/a-data-driven-exploration-of-crime-trends-in-los-angeles-6124c2980eda

Niklas Frechen

Center for Advanced Internet Studies (CAIS)

E-Mail niklas.frechen@cais-research.de

Niklas Frechen is a doctoral researcher at the Center for Advanced Internet Studies (CAIS) in Bochum, Germany. He has a background in cognitive psychology and linguistics and is currently focusing on the need for analog in a digitalized world.

Pauline Heger

Center for Advanced Internet Studies (CAIS)

E-Mail pauline.heger@cais-research.de

Pauline Heger is a doctoral researcher at the Center for Advanced Internet Studies (CAIS) in Bochum, Germany, with a background in communication science. Her research focuses on social diffusion of innovation, sustainable technologies, and HCI.

Christoph Bieber

Center for Advanced Internet Studies (CAIS)

E-Mail christoph.bieber@cais-research.de

Christoph Bieber is a research professor at the Center for Advanced Internet Studies (CAIS) and a political scientist at the University of Duisburg-Essen. He leads the research program Digital Democratic Innovations.

Mennatullah Hendawy

Center for Advanced Internet Studies (CAIS)

E-Mail mennatullah.hendawy@cais-research.de

Mennatullah Hendawy is an interdisciplinary urban planner working on cities and technology towards equity and sustainability. She is a Postdoc at the Center for Advanced Internet Studies (CAIS) in Bochum, Germany, and an Assistant Professor at Ain Shams University in Cairo, Egypt.