DEVELOPING A PRIVACY-AWARE MAP-BASED CROSS-PLATFORM SOCIAL MEDIA DASHBOARD FOR MUNICIPAL DECISION-MAKING

D. Weckmüller*, A. Dunkel

Institute of Cartography, Technische Universität Dresden, Helmholtzstr. 10, 01062 Dresden, Germany
dominik.weckmueller@mailbox.tu-dresden.de; alexander.dunkel@tu-dresden.de

Commission IV, WG IV/4

KEY WORDS: social media, privacy, spatial dashboard, HyperLogLog, urban planning

ABSTRACT:

Social media (SM) has become a principal information source and the vast amounts of generated data are increasingly used to inform various disciplines. Most platforms, such as Instagram, Twitter or Flickr, offer the option to tag a post with a location or a precise coordinate, which also fosters applications of data in the geospatial fields. Notwithstanding the many ways in which these data could be analyzed and applied, scandals such as Cambridge Analytica have also shown the risks to user privacy that seem inherently part of the data.
Is it possible to mitigate these risks, while maintaining the collective usability of this data for society questions? We identify urban planning as a key field for socio-spatial justice and propose an open source map-based cross-platform dashboard, fueled by geospatial SM, as a supporting tool for municipal decision-makers and citizens alike. As a core part of this tool, we implement a novel privacy-aware data structure that allows for both, a more transparent, encompassing data ground for municipalities, and a reduced data collection footprint, preventing the misuse of data or compromising user privacy.

1. INTRODUCTION

Location-based SM networks (LBSN) have produced an unprecedented data base over the past 15 years. According to Ilieva and McPhearson (2018), “the enormous scale and timely observation are unique advantages”. They hold “enormous opportunities for communication research” (Hoffmann, Heft, 2020) that can be applied to various fields, such as urban planning.
For example, as one of the more popular LBSNs, Instagram is increasingly encouraging the indication of location when creating content. The data are mostly public, relatively easy to query, while scientific analysis methods, such as natural language processing based on artificial intelligence (AI), are well-researched. Albeit of these advances in research, practical application still lacks behind. The reasons are manifold. State-of-the-art LBSN research tends to focus on ever evolving algorithms providing expert knowledge. Speaking through the “Data Information Knowledge Wisdom (DIKW) Hierarchy” by Cooley (1987), Zeleny (1987) and Ackoff (1989), this widens the gap between a small, highly skilled scientific community reaching knowledge and wisdom about data, and the majority of the population lacking the skills and appropriate platforms to even access the data. In addition, immature ethical and legal guidelines complicate the use of original data outside research (Heikinheimo et al., 2020; Fuller, 2019).
While research in urban planning has consistently identified a “data gap” (Voskamp et al., 2018) we identify a particular platform gap in urban planning. Urban planners, particularly policy makers are still lacking an encompassing knowledge ground. The urge of such is ever-increasing, as nowadays most of the world’s population is already living in urban areas with the share continuously growing (United Nations, Department of Economic and Social Affairs, Population Division, 2019), while socio-spatial disparities are rising.
Our stated research goal is therefore to provide citizens, laypersons and municipal decision-makers with an easily accessible and transparent LBSN Dashboard, as an open source platform to supplement spatial decision-making processes in municipalities. The prototypical implementation of a HyperLogLog (HLL) based LBSN data structure, leaning on Flajolet et al. (2007) and Dunkel, Löchner and Burghardt (2020), is therefore proposed for privacy-aware analysis and used as a backend for the map-based dashboard.

2. URBAN PLANNING, SPATIAL EQUALITY AND EQUITY

Urban planning and management, within the scope of this study, are defined in a broad sense as “technical, social and political process concerned with the design, development and maintenance of land use in an urban environment” (van Maarseveen, Martinez, Flacke, 2019). As such, “[u]rban planning encompasses activities such as strategic thinking, research and analysis, public consultation, urban design and policy implementation“ (ibid.), which requires a solid data ground for informed decision making. The overarching goal is to steer against uncontrolled urbanization, strategically counteracting “multiple forms of inequalities, exclusion and deprivation” (ibid.).
The underlying ruleset for urban planners is provided by laws, with equality and equity being the fundamental principles. In essence, equality and equity are about people being able to participate in society in a similar way, or being entitled an equal and fair access to certain resources.
Embracing the idea of equity through spatially just urban planning, requires allocation in the sense of maximizing social

* Corresponding author

justice (Harvey, 2010), which in turn requires to address the individual "variations in demands and needs in the population".
For this, it is theoretically necessary to have "information concerning the utility scales of each individual in the population" (ibid.). This is difficult to attain fully, but emphasizes the necessity and indispensability of spatial information (ibid.).
In conclusion, spatial information plays an important role in enabling a fair and livable city, by ensuring evidence-based decisions and policies based on transparent democratic negotiation processes. An LBSN dashboard could contribute to fill the gap, by providing access to missing information about people's individual preferences and needs. Similar to the "Spatial Decision Support Systems" proposed by Yeh (1999), it could serve as a data hub to strengthen the integration of spatial equality and spatial equity in daily decision-making processes.

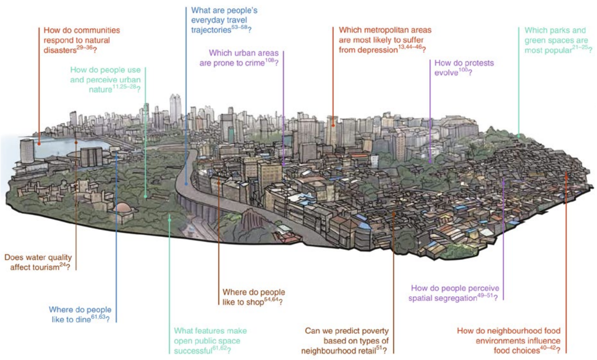

Figure 1. The wide range of emerging opportunities for urban-sustainability research provided by big data from social media (Ilieva, McPhearson, 2018).

Apart from social equity, Ilieva and McPhearson (2018, Fig. 1) identify four more core fields for SM big data application: environmental sustainability, public health, social equity, mobility and economic development, all of which are deeply entrenched with urban planning and management.
However, even though „[i]nteractive online dashboards are an accessible way to summarize complex information to the public" (Pellert et al., 2020), few dashboard-like tools exist that make use of LBSN. Prototypes exist in the areas of health (e.g., Pellert et al., 2020; Padmanabhan, 2014), in a disaster context of storms and floods (e.g., Tsou et al., 2015) and in a geomarketing context for site analysis (e.g., Anderson et al., 2019; Lin et al., 2016). Still, these are not publicly available. As boyd and Crawford (2012) emphasize, "[w]rangling APIs, scraping, and analyzing big swathes of data is a skill set generally restricted to those with a computational background […]" and should therefore be made more accessible. At the time writing, there is no publicly available SM dashboard that allows to study privacy-aware LBSN data from different platforms, bundled together and made accessible to laypersons for different contexts while taking concerns about privacy, law and ethics seriously. Ilieva and McPhearson (2018) point out the increasing need for such an attempt:
„Global urban science remains fragmented and disconnected from global and local policy and planning, highlighting the need for new tools and data to advance understanding of complex urban dynamics, and to support decision-making for sustainability transformations."

3. DATA PRIVACY

Data privacy developed at a rapid pace in the 2000s. It used to be a rather vague concept, despite the fact that it has become part of most people's everyday lives (Barker et al., 2009).
Generally, as summarized by Moore (2008), literature agrees on common aspects of privacy, that is the right to decide autonomously whom to share personal data with and who might know about one's actions. The default should always be beneficial for the data owner, meaning no information is available to third parties. Contrary to a missing clear-cut definition, the data privacy goal within the scope of this thesis is pretty clear, namely to "release statistical information about the population who have contributed to the data without breaching their individual privacy" (Gehrke, Lui, Pass, 2011).
In recent literature, the conflict between user privacy and growing commercial interests is elaborated. Li, Sharma and Mohanty (2020) claim that disclosing data can cause serious harm to individuals, yet "driven by economic advantages" companies capture, store and use data from many different sources and turn these to knowledge through analysis. Since knowledge means profit, LBSN companies have a high incentive of storing, aggregating and collecting more data about their users. The "frequent incidents of personal data leakage" (ibid.) such as the scandal about Cambridge Analytica prove the controversy of such commercial practices. As Baik (2020) summarizes, a general shift can be observed from privacy as universal right or dignity towards a good or a commodity. The problem here is that "the creator of data – a user – is not necessarily equal to the monetary beneficiary of the data – a digital platform" (ibid.).
At this point, a common possible misconception is that privacy is automatically given if the individual data are only encrypted well enough and are not accessible to the public. Instead, Sweeney (2002) claims "computer security is not privacy protection". Privacy does not end with authentication and access control, but begins with the data themselves (ibid.). Put simply, privacy must be granted to the data owner at the stage where a third party has access to the data or in other words: "anything that can be learned about a respondent from the statistical database should be learnable without access to the database" (Dalenius, 1977, as cited in De Capitani Di Vimercati et al., 2012).
In this sense, the strictest of definitions of privacy is called differential privacy (DP). DP provides guarantees to user privacy by ensuring "that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis" (Dwork, 2008). It is usually achieved by adding noise to the data, which, according to Dunkel, Löchner and Burghardt (2020), provides certain limits to usability in practice, particularly to streaming data such as LBSN, where the characteristics of the entire dataset are not known ahead of time. Their idea follows a different, more flexible approach, based on the data abstraction algorithm HyperLogLog (HLL), in addition to classic computer security approaches including anonymization and encryption. Through various parameters, it is possible to fine tune the balance between individual privacy and analytical freedom in different analysis contexts.
Particularly in urban planning, geoprivacy, as a subcategory of data privacy, plays an important role. As Kessler and McKenzie (2018) see it, "[i]nformation about an individual's location is substantially different from other kinds of personally identifiable information" as they allow "for a broad range of location-based inferences, such as information about their health, consumer behavior, or social status". Despite a high general concern about privacy matters in population (ibid.), there seems to be little consciousness about geoprivacy in everyday life, for instance, when sharing locations with services and apps (ibid.). A recent study of Chen et al. (2021) shows "[…] that there is no relationship between privacy concern […] and the number of data-sharing authorizations, confirming the puzzling data privacy paradox". According to the authors the reason is "economic benefits of sharing personal data with mini-programs" (ibid.). Unfortunately, such a loose laissez-faire attitude is a trend that

can also be observed in science. Notwithstanding broad privacy discussions, "unmasked confidential [geo-] data increased" in scientific publications (Kounadi, Leitner, 2014). According to Kessler and McKenzie (2018), geoprivacy is a complex tension field between societal developments, useful services fueled by user data, legal and ethical considerations making it "difficult to tackle as a whole" (ibid.).
In conclusion for application to urban planning, there are different privacy–utility tradeoffs to be considered. On one end of the continuum is the need of society for generalized information, to plan ahead and improve living together in cities. On the other end are the privacy needs of individuals, to withdraw generally or in certain situations in order to protect personal information. The vision that is taking shape is that analysts, such as urban planners, and citizens, collaborate to improve the use and design of public spaces and work together towards a fair allocation of resources in cities.

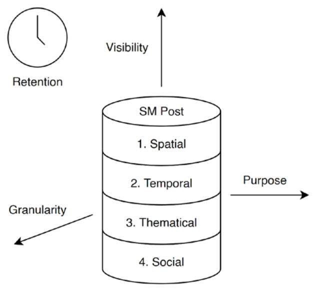

Figure 2. Post facets and privacy dimensions (modified after Löchner, Dunkel, Burghardt, 2018; Barker, 2009).

To summarize, Figure 2 provides an overview of the various starting points for LBSN privacy measures, which is to be based on the four dimensions of privacy according to Barker (2009) and the four facets according to Löchner, Dunkel and Burghardt (2018) centered around LBSN data. How exactly post facets and dimensions can contribute to data privacy is further described in chapter 4.2.
As working with sensitive data generally is a delicate matter it requires not only prudence but also a profound discussion of potential risks, ethical and legal implications. Unfortunately, as Heikinheimo et al. (2020) put it, "legislation and ethical guidelines on using publicly but passively contributed data sets in research are still immature" due to its novelty.

4. METHODS AND DATA

The dashboard prototype consists of several individual components that are presented in the following chapters. Leaning on the actual data workflow, every section deals with one fundamental backend component, as described in Figure 3.
Privacy is gradually increased across the whole process. The workflow can be summarized as follows. Starting from different LBSN such as Instagram, Twitter or Flickr, the relevant posts are queried. For this process, either the respective APIs or data scraping can be used. Importantly, when data is retrieved, it is not stored directly. Instead, information, such as posts, are split up into their so-called 'bases' in-memory and partially recomposed. These bases could be individual dates, times, locations, or semantics. For particularly sensible information, such as user identifiers (UIDs), Dunkel et al. (2019) propose the use of HLL, only storing approximate summaries of data. For every base, an individual HLL set is created in the database (see chapter 4.2 and 4.3), ready to be queried by the frontend or API user (see chapter 4). Generalization and filtering techniques are applied ad-hoc,

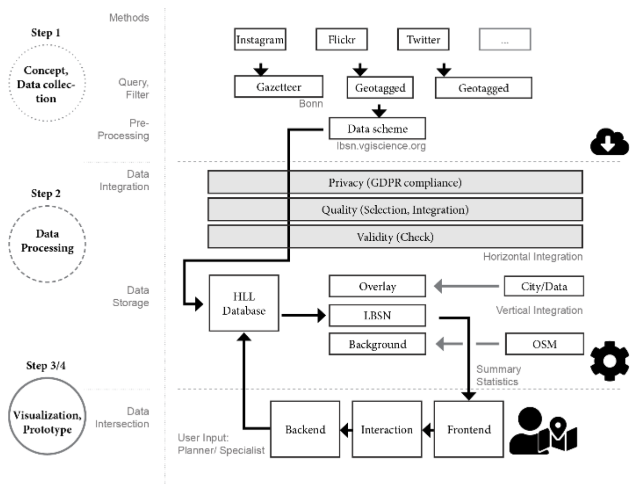

Figure 3. Dashboard methodology.

depending on the particular context, to reduce granularity of information to the maximum required resolution. By using the unique capabilities of stored HLL sets, it is possible to apply further generalization afterwards, during retrieval and display of information (see chapter 4.2 - 4.4). The detailed technical open source stack (PostgreSQL database with PostGIS extension and fastAPI web framework) is further described in the respective online repositories (gitlab.vgiscience.de/lbsn/databases/hlldb; github.com/do-me/LBSN-Dashboard).

4.1 Data Sources

There is a growing and ever-changing range of LBSN (Fig. 3, Step 1). For the dashboard prototype described in this study, data from the LBSN Instagram, Twitter and Flickr are used for reasons to be clarified below. Each LBSN as well as the respective access to the data are briefly presented, followed by other data sources which can be currently processed in the dashboard prototype.
According to official internal statistics of Meta (2021, as of 05/2021), Instagram has the largest user base amongst LBSN with >500 million daily active users (DAUs) (as of 09/2017) or >1 billion monthly active users (as of 06/2018). Amongst LBSN it has a relatively high percentage of geotagged posts with so-called "Facebook locations" having a certain description and a coordinate.
Unfortunately, as Meta seldom reveals official and precise statistics, the geotag quota cannot be quantified exactly. Empirical studies have found that the percentage may range roughly from 21-25% (Boy, Uitermark, 2015) to 41% (Fiallos et al., 2018). Still, these studies must be interpreted with caution as their samples are very small with respect to Instagram's entire database.
Twitter is another popular social network with a microblogging service including a geotagging function. One can post so-called tweets, containing images, text and tags limited to a maximum of 140 characters (Morstatter et al., 2013). Lastly, Twitter (2022a) reported 214.7 million monetizable DAUs.
Geotagging in Twitter is either based on POIs of the company foursquare or can contain a custom location label and the exact GPS-Coordinate, but only when using Twitter's mobile app (Twitter, 2021b). For the posts being tagged with a GPS-Coordinate, the precision in comparison to Instagram is higher and adds further detail.
Few posts on Twitter, also called "Tweets", are geotagged. In fact, some studies talk of around 1% (Morstatter et al., 2013),

https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-545-2022

others come to a conclusion of around 2% excluding a varying percentage which can be approximately identified through time zones and text matching (Burton et al. 2012) or 2.3% in a large-scale study analyzing 40 billion tweets (Huang, Carley 2019). Flickr is an "online photo management and sharing application" where members can connect and interact (Flickr, 2022a). It hosts more than 10 billion photos (ibid., 2015, as of 2015) partially with metadata such as timestamp and coordinates and counts around 60 million visitors every month (ibid., 2022b). As of 2009 roughly 3.3% of all Flickr photos were geotagged (Flickr, 2009). Unlike Facebook or Instagram, Flickr provides a limited feature to search photos via a map interface (see Flickr, 2021c). In SM research Flickr it is well known for its open API which is freely accessible for everyone for non-commercial use (ibid.). Instagram and Twitter instead have a larger data potential, offering a strictly limited API to researchers while cooperating with partner companies who have direct access to larger amounts of data.
Generally, all three platforms have a public posting policy, so that theoretically most of the posts are publicly available even without login. Other LBSN, such as YouTube, TikTok or future platforms, can easily be integrated.
Besides LBSN, open data are a fundamental data source. The Open Knowledge Foundation (2022) defines open knowledge as "any content, information or data that people are free to use, re-use and redistribute — without any legal, technological or social restriction.". Most municipalities, regions and countries have their own open data portals with specific open data sets that can be used for any purpose (e.g., Bundesstadt Bonn, 2022). Amongst open data, OpenStreetMap takes a special place as it is mapping most parts of our built and natural environment such as "roads, trails, cafés [or] railway stations" (OpenStreetMap Foundation, 2022). It is a great source for base maps, points of interest, borders or other geometries for further processing, overlays or intersections. If needed, commercial data can serve as additional data sources whenever higher accuracy or very specific topics are investigated.

4.2 Data Processing

As different platforms each have different API structures and data formats, the dashboard backend requires a common data structure to make each base comparable across all platforms (Fig. 3, Step 2). Dunkel et al. (2019) provide such a conceptual LBSN framework, in the context of reactions to events, based on which SM posts can be classified into different categories for analysis purposes. They divide a single SM post into four so-called facets (ibid.):

- Temporal: Time or period
- Spatial: Location or spatial relation
- Social: Demographic makeup
- Thematic/Topical: Meaning and content

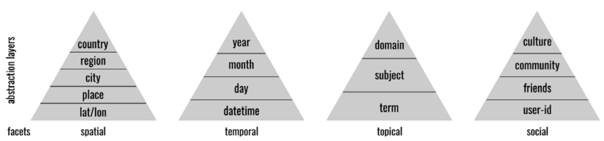

Figure 4. Abstraction layers for each facet (Löchner, Dunkel, Burghardt, 2018).

The facets can be divided into different granularities (Fig. 4) and are used in this study for three reasons. First, it provides an analysis scheme, proven in the literature in similar studies to examine individual posts more closely (ibid.). Second, it allows different LBSNs and their posts to be compared and brought to a common level of analysis, which is an important factor in the context of an SM dashboard. Third, there is an already existing practical implementation ('LBSN Structure', lbsn.vgiscience.org), offering an easy Docker-based setup as well as a solid database for the application backend.

4.3 Data Integration

The dashboard is designed for privacy-aware live usage and hence must not only adhere to the strict guidelines of the European Union General Data Protection Regulation (GDPR, 2016), but ensure that it is technically impossible to reproduce original posts, even under full public disclosure of the database. For this purpose, certain subsets can be created in advance. For example, if one always wanted to monitor all data of the previous day, these would be read into a separate HLL set. Retroactively, these can have not only arbitrary temporal filters, but also spatial filters, by removing subsets of the original data before creating the set. Currently speaking, with respect to the four SM facets (Fig. 4), the first two (temporal, spatial) are already covered in the prototype.
The social facet can be roughly covered by the selection of the LBSN, since each LBSN has a different user base. Furthermore, simple filters (e.g., only male users over 30) could be pre-grouped and studied in combination with other attributes, as long as the original data allows such a high resolution. Such base combinations can also be generated with locations, coordinates or geohashes. However, the use of coordinates only makes sense if they occur more than once. On Flickr, for example, individual coordinates are often stored for posts, whereas on Instagram only the location IDs are stored causing a coarser resolution. It is therefore more productive to generalize spatial information ahead to grids, shapes, or official boundaries.
In contrast to the other facets, the procedure for the manifold thematic facet requires to break down the content of posts into 'atomic' bases, such as terms, hashtags, or emoji (Dunkel, Löchner, Burghardt, 2020). Using this approach, the connection between UID and post content can be dissolved, as this link is highly problematic from a privacy perspective (ibid.). This procedure not only generates a large part of privacy, but is also the foundation for HLL.

👩 SunLover22 📅 30/07/2021 🕑 02:15 P.M. 📍 Capo di Conca, SA, Italy Today enjoying the ***wonderful*** ☀️ on the 🏖️ 🌊 🌴 #costaamalfi with @lovingsister #siblings # sun #beach #summer 😊 [Photo with accessibility caption: "This image could contain 2 persons on the beach"]

Table 1. Fictive sample SM post with metadata.

Table 1 illustrates an example post that is broken down into its bases in Table 2, consisting of metadata such as UIDs, timestamp and location, the post content consisting of text, emojis, hashtags, user tags and other photo metadata.
All this information is thus separated from each other, that is each piece of information is read into a different HLL set (Fig. 5).
If there is no HLL set for the respective element, a new one is created, leading to a many different HLL sets with different sizes. For instance, the HLL set for the hashtag "#costaamalfi" (Table 1/2) would now have a cardinality (size) of approximately 1 if no

other post used the same hashtag before.
Since larger HLL sets usually provide larger benefits to privacy (Desfontaines et al., 2019), filtering based on thresholds, or blocklists and allowlists should be used.

Category	Post Base	Sample
Post Metadata	UID Timestamp Location	SunLover22 30/07/2021, 02:15 P.M Capo di Conca, SA, Italy
Post Content	Terms Emojis Hashtags User Tags	“Today”, “enjoying”, “the”, “wonderful”, “on”, “the” “with” “🌞“, “⛱”, “🌊”, “🌴”, “😊” “#costaamalfi”, “#siblings”, “#sun”, “beach”, “summer” “@lovingsister”
Photo Metadata	Photo Accessibility Caption	[A photo of two persons on the beach] “2 persons”, “on the beach”

Table 2. A fictive SM post split into its atomic bases.

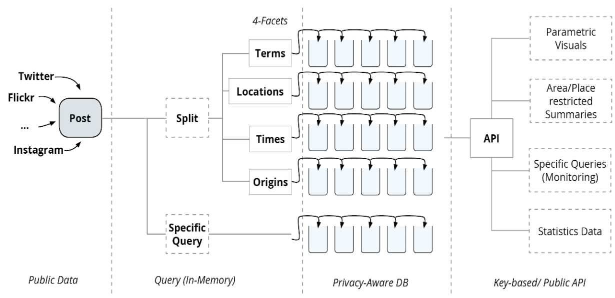

Figure 5. HLL data workflow for splitting SM posts in its bases.

4.4 Data Exploration

Adhering to the described framework and Figures 2-5, LBSN data are only permanently stored as bases. These bases, either atomic or recombined, can then be put back together for interactive exploration, in a limited way based on the capabilities of HLL, such as aggregating, intersecting, or interlacing with different data sets, as described in chapter 4.1. For instance, analysts interested in particular day or night population behavior, could use a custom time filter applied to a recombined base of latitude/longitude, user ID and daytime. Or, using either OSM or a dedicated municipal land use plan for custom area of interest geometry filters, one can filter for particular areas in a city, such as urban green spaces, using the recombined HLL base of hashed latitude/longitude and approximate UIDs. Lastly, a certain topic in a particular area could be explored through a recombination of latitude/longitude, UIDs and terms.
More abstract data exploration is possible, such as combinations of weather data with latitude/longitude, user ID and date. The myriad of combination possibilities offers a great variety of insight to different applications.

5. APP

5.1 Frontend

All of the previously described options of exploration require taking into account the analyst, which is done through interactions in the frontend of the web application (Fig. 3, Step 3/4). There is a range of functions already implemented in the current dashboard prototype for the city of Bonn, Germany (geo.rocks/dashboards/bonn) as well as for Waynesboro, VA, USA (geo.rocks/dashboards/waynesboro). Respective video tutorials can be found on the demo pages.
The whole data flow (Fig. 3, 5) is implemented for the previously described LBSN platforms. The status quo implementation uses requests for the entire HLL-DB by default, meaning aggregation of all LBSN. In this manner an additional layer of privacy is introduced, as the client cannot know what share originates from individual LBSN.

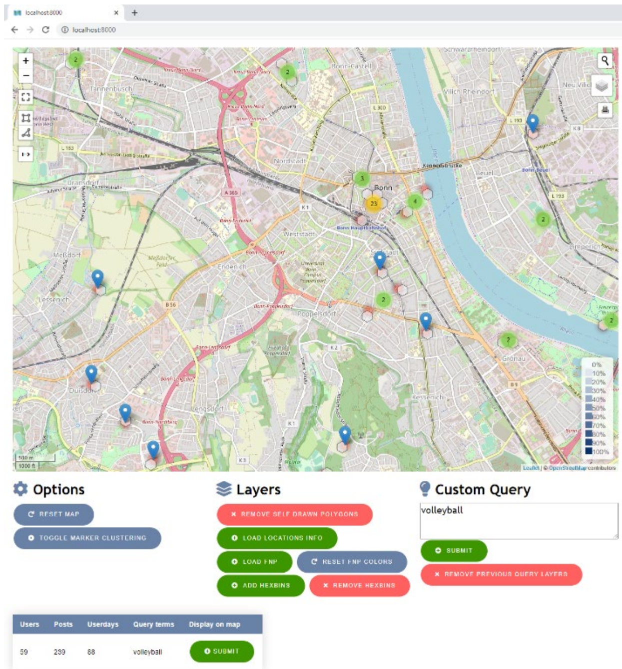

Figure 6. Dashboard frontend interface.

The two already implemented core functionalities are spatial and thematic queries, which could also be combined. The spatial query can be performed by drawing a polygon, a multi-polygon or a multi-polygon with holes (islands) on a Leaflet-based map through the Leaflet-Geoman plugin (Fig. 6, top left). The thematic query can be performed by searching for certain keywords (terms) that occur in post captions (Fig. 6, bottom right). In this way, it is possible to explore hotspots for thematic interests of various kinds.

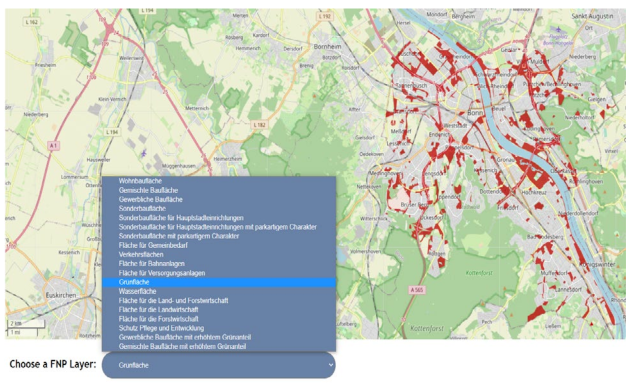

Figure 7: Dashboard plugin for choosing land use category of Bonn.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-545-2022 | © Author(s) 2022. CC BY 4.0 License.

The request should happen at least in English and the local language (DE in the Bonn context), since SM generally tend to multilingualism with other minority languages and a high affinity for "'intra-linguistic' variety of register and styles" as shown in a case study by Leppänen and Kytölä (2017). As mentioned previously, an interface for adjusting the temporal and social facet is yet to be implemented.
The front- and backend are built in a way that it is easy to add 'plugins'. As can be seen in Figure 7, such a plugin is provided for the Bonn use case. The button "LOAD FNP" (Fig. 6) adds the land-use plan (DE: "Flächennutzungsplan" or "FNP") and provides further analysis possibilities. A category such as "Green Spaces" (DE: "Grünfläche") can be selected. The geometries are then displayed and its metrics analyzed, similar to what is shown in Figure 6. In the same way, any other geometries could be queried, which opens a broad range of application possibilities. For example, public participation could be encouraged via the same interface, for instance, by allowing public comments on locations or including already existing citizen petitions.

5.2 Interactivity and Participation: A Municipal SM platform

The city of the future depends on data as it depends on active citizen participation. Here, the status quo that citizen data is shared and used through opaque SM APIs and hosted elsewhere, is both hard to accept as it appears futile to change (Sunstein, 2019). While a dashboard, such as the one proposed here, may provide guarantees to user privacy, trust must be gained gradually. In this sense, active communication and transparency are key aspects to be considered when the dashboard is actually put to use. Importantly, the proposed bases are not bound to SM. Given the technical advances and pervasiveness of mobile devices, it appears at least possible that users and analysts collaborate more closely, without an intermediate third party (e.g., SM).
The future development could therefore focus on transition features, allowing a shift from passive towards more active modes of communication (compare Ghermandi, Sinclair, 2019). For instance, hashtag campaigns may be a first step to motivate users on various SM to contribute data more explicitly, for example, in the context of critical urban planning contexts. Here, the dashboard would need to be advanced by allowing analysts to add new campaigns and explore the results of existing ones. Further ahead, integration of qualitative content may provide an additional layer, next to quantitative base statistics from global SM. Explicitly adding photos or descriptions on a map, would allow citizens to better communicate opinions or needs, in a more detailed fashion, while being aware of the implications to privacy. The vision that is taking shape is that analysts, such as urban planners, and citizens, collaborate to improve the use and design of public spaces and towards a fair allocation of resources in cities.

6. DISCUSSION AND CONCLUSION

There is a lively debate in literature about equality and equity, not only because they are "widely confused" (Bronfenbrenner, 1973). Equality is "basically objective" as the amount of a certain good per person can be measured whereas equity is "basically subjective" as it is subject to "ethical judgment" (ibid.).
When defining equity more towards the concrete spatially relevant need of a person, it becomes obvious that certain resources are more in demand by certain groups and less by others. This idea is illustrated well by a simple example, because "[...] equidistant access to kindergartens is not useful when the population's age distribution is considered. Not every district in a city has a similar need for kindergartens" (Dunkel, 2016).
Soja (2013) builds on this concept and speaks of "spatial justice" or "spatial injustice [...] as the outcome of countless decisions made about emplacement, where things are put in space". Soja (ibid.) links the production of spatial injustice to the "right to the city" and claims that many institutions such as the European Union already exist to specifically combat spatial injustice. However, these institutions in turn require spatial information that is specifically processed. In order to counter the "discriminatory geographies", capacities, that is "geographical information systems" (ibid.), have already been used in the past. Still, such systems can only promote spatial justice in combination with "will and awareness" (ibid.). As for example "socio-economic and ethnic divides are entrenched and socially reproduced through physical space and everyday life" very much in a Foucauldian sense, SM data are "a useful asset for social-equity research in cities" which GIS researchers already have investigated (Ilieva, McPhearson, 2018).
The discourse around Spatial (In-) Justice (compare Soja, 2013), the debate of (in-) equality as well as (in-) equity and its continuous reproduction (Amis, Mair and Munir 2020) consists an urgently needed debate to understanding these socio-spatial realities that seem not to have been improved by more data or more sophisticated algorithms alone.
In fact, amongst many, Sloane and Moss (2019) title it as "AI's social sciences deficit" while Leslie et al. (2021) go even further and claim that these algorithms might yet be another perpetuating and amplifying factor for growing inequality and inequity and, besides, raising important questions about privacy and data sovereignty.
While SM are not universally representative, they are to be considered as supplementary data source. Herein, the dashboard is increasing representativeness by providing access to a more holistic view of interests, needs, and preferences of different people and groups of people who partially even might have had little to no direct voice in planning processes so far. In addition, the static SM data can be supplemented with actively contributed, qualitative content in the future for a more dynamic and participative communication.
Too few research attempts were made to enable laypersons to apply the highly needed spatial knowledge to their particular domain – especially in urban planning. Put simply: "No matter how sophisticated and advanced it is, a decision support system is useless if it is not being used by decision-makers" (Yeh, 1999). In the balancing act of maximizing collective benefits while reducing impairments of individual's privacy, the combination of the estimative nature of HLL, the LBSN framework, a secure backend and the accessible frontend developed here provide a privacy-aware basis that should be used for municipal decision-making in urban planning and for increased decision transparency for citizens.

REFERENCES

Ackoff, R. L., 1989. From data to wisdom. *Journal of applied systems analysis*, *16*(1), 3-9.

Amis, J. M., Mair, J., Munir, K. A., 2020. The organizational reproduction of inequality. *Academy of Management Annals*, 14(1), 195-230.

Anderson, J., Casas Saez, G., Anderson, K., Palen, L., Morss, R., 2019. Incorporating context and location into social media analysis: A scalable, cloud-based approach for more powerful data science. In T. X. Bui (Ed.) *Proceedings of the 52nd Hawaii*

International Conference on System Sciences (pp. 2274- 2283). Hawaii International Conference on System Sciences.

Baik, J. S., 2020. Data privacy against innovation or against discrimination?: The case of the California Consumer Privacy Act (CCPA). *Telematics and Informatics*, 52, 1-10.

Barker, K., Askari, M., Banerjee, M., Ghazinour, K., Mackas, B., Majedi, M., Pun, S. Williams, A., 2009. A data privacy taxonomy. In A. P. Sexton (Ed.) Dataspace: The Final Frontier – *26th British National Conference on Databases*, BNCOD 26 (pp. 42-54). Springer.

Boy, J. D., Uitermark, J., 2015. Capture and share the city: Mapping Instagram's uneven geography in Amsterdam. In The Ideal City: between myth and reality – Representations, policies, contradictions and challenges for tomorrow's urban life Urbino (Vol. 27) (pp. 1-20). Research Committee 21 (RC21).

boyd, D., Crawford, K., 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15(5), 662-679.

Bronfenbrenner, M., 1973. Equality and equity. *The ANNALS of the American Academy of Political and Social Science*, 409(1), 9-23.

Bundesstadt Bonn, 2022. Offene Daten Bonn. opendata.bonn.de.

Burton, S. H., Tanner, K. W., Giraud-Carrier, C. G., West, J. H., Barnes, M. D., 2012. "Right time, right place". Health communication on Twitter: value and accuracy of location information. *Journal of medical Internet research*, 14(6), e2121.

Chen, L., Huang, Y., Ouyang, S., Xiong, W., 2021. The Data Privacy Paradox and Digital Demand. National Bureau of Economic Research Working Paper, 28854, 1-53.

Cooley, M., 1987. *Architect or bee? The human price of technology* (Second). The Hogarth Press.

Dalenius, T., 1977. Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15(5), 429-444.

De Capitani Di Vimercati, S., Foresti, S., Livraga, G., Samarati, P., 2012. Data privacy: definitions and techniques. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 20(06), 793-817.

Desfontaines, D., Lochbihler, A., Basin, D., 2019. Cardinality Estimators do not Preserve Privacy. *Proceedings on Privacy Enhancing Technologies*, 2019(2), 26-46.

Dunkel, A., 2016. Assessing the perceived environment through crowdsourced spatial photo content for application to the fields of landscape and urban planning. [Doctoral dissertation, TU Dresden]. CORE open access research.

Dunkel, A., Andrienko, G., Andrienko, N., Burghardt, D., Hauthal, E., Purves, R., 2019. A conceptual framework for studying collective reactions to events in location-based social media. *International Journal of Geographical Information Science*, 33(4), 780-804.

Dunkel, A., Löchner, M., Burghardt, D., 2020. Privacy-Aware Visualization of Volunteered Geographic Information (VGI) to Analyze Spatial Activity: A Benchmark Implementation. *ISPRS International Journal of Geo-Information*, 9(10), 1-21.

Dwork, C., 2008. Differential privacy: A survey of results. In M. Agrawal, D.-Z. Du, Z. Duan, A. Li, (Eds.) *International conference on theory and applications of models of computation* (pp. 1-19). Springer.

European Union General Data Protection Regulation (GDPR), 2016: Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

Fiallos, A., Jimenes, K., Fiallos, C., Figueroa, S., 2018. Detecting topics and locations on Instagram photos. In L. Terán, A. Meier (Eds.) 2018 International Conference on eDemocracy & eGovernment (ICEDEG) (pp. 246-250). The Institute of Electrical and Electronics Engineers.

Flajolet, P., Fusy, É., Gandouet, O., Meunier, F. 2007. Hyperloglog: the analysis of a nearoptimal cardinality estimation algorithm. *Analysis of Algorithms 2007* (AofA07), 127–146.

Flickr, 2022a. The Flickr Developer Guide. flickr.com/services/developer.

Flickr, 2022b. Work at Flickr. flickr.com/jobs.

Flickr, 2022c. Flickr map. flickr.com/map.

Flickr, 2015. Find every photo with Flickr's new unified search experience. blog.flickr.net/en/2015/05/07/flickr-unified-search.

Flickr, 2009. 100,000,000 geotagged photos (plus). code.flickr.net/2009/02/04/100000000-geotagged-photos-plus.

Fuller, M., 2019. Big data and the facebook scandal: Issues and responses. *Theology*, 122, 14–21.

Gehrke, J., Lui, E., Pass, R., 2011. Towards privacy for social networks: A zero-knowledge based definition of privacy. In Y. Ishai (Ed.) *Theory of cryptography conference* (pp. 432-449). Springer.

Ghermandi, A., Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: A systematic map. *Global environmental change*, *55*, 36-47.

Harvey, D., 2010. *Social justice and the city* (Vol. 1). University of Georgia Press.

Heikinheimo, V., Tenkanen, H., Bergroth, C., Järv, O., Hiippala, T., Toivonen, T., 2020. Understanding the use of urban green spaces from user-generated geographic information. *Landscape and Urban Planning*, 201, 103845.

Hoffmann, M., Heft, A., 2020. "Here, There and Everywhere": Classifying Location Information in Social Media Data-Possibilities and Limitations. *Communication Methods and Measures*, 14(3), 184-203.

Huang, B., Carley, K. M., 2019. A large-scale empirical study of geotagging behavior on twitter. In F. Spezzano, W. Chen, X. Xiao (Eds.) *Proceedings of the 2019 IEEE/ACM International*

Conference on Advances in Social Networks Analysis and Mining (pp. 365-373). Institute of Electrical and Electronics Engineers.

Ilieva, R. T., McPhearson, T., 2018. Social-media data for urban sustainability. *Nature Sustainability*, 1(10), 553-565.

Kessler, C., McKenzie, G., 2018. A geoprivacy manifesto. *Transactions in GIS*, 22(1), 3-19.

Kounadi, O., Leitner, M., 2014. Why does geoprivacy matter? The scientific publication of confidential data presented on maps. *Journal of Empirical Research on Human Research Ethics*, 9(4), 34-45.

Leppänen, S., Kytölä, S., 2017. Investigating multilingualism and multisemioticity as communicative resources in social media. In M. Martin-Jones, D. Martin (Eds.) Researching multilingualism: Critical and ethnographic perspectives (pp. 155–171). Routledge.

Leslie, D., Mazumder, A., Peppin, A., Wolters, M. K., Hagerty, A., 2021. Does "AI" stand for augmenting inequality in the era of covid-19 healthcare?. *The bmj*, 372.

Li, Z., Sharma, V., Mohanty, S. P., 2020. Preserving data privacy via federated learning: Challenges and solutions. *Consumer Electronics Magazine*, 9(3), 8-16.

Lin, J., Oentaryo, R., Lim, E. P., Vu, C., Vu, A., Kwee, A., 2016. Where is the goldmine? Finding promising business locations through Facebook data analytics. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media* (pp. 93-102). Association for Computing Machinery.

Löchner, M., Dunkel, A., Burghardt, D. (2018). *A privacy-aware model to process data from location-based social media.* In D. Burghardt, S. Chen, G. Andrienko, Gennady, N. Andrienko, R. Purves, & A. Diehl (Eds.) VGI Geovisual Analytics Workshop, colocated with BDVA 2018 (pp. 1-5). The Institute of Electrical and Electronics Engineers.

Meta, 2021. Facebook for Business – Tell your brand story your way with Instagram. web.archive.org/web/20220118191633/https:/www.facebook.com/business/marketing/instagram.

Moore, A. D., 2008. Defining privacy. *Journal of Social Philosophy*, 39(3), 411-428.

Morstatter, F., Pfeffer, J., Liu, H., Carley, K., 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1) (pp. 400-408). The AAAI Press.

Open Knowledge Foundation, 2022. What is open?. okfn.org/opendata.

OpenStreetMap Foundation, 2022. OpenStreetMap provides map data for thousands of websites, mobile apps, and hardware devices. openstreetmap.org/about.

Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K., Liu, Y., 2014. FluMapper: A cyberGIS application for interactive analysis of massive location-based social media. *Concurrency and computation: Practice and Experience*, 26(13), 2253-2265.

Pellert, M., Lasser, J., Metzler, H., Garcia, D., 2020. *Dashboard of sentiment in Austrian social media during COVID-19*. arXiv preprint. arxiv.org/pdf/2006.11158.pdf, 1-23.

Sloane, M., Moss, E., 2019. AI's social sciences deficit. *Nature Machine Intelligence*, 1(8), 330-331.

Soja, E. W., 2013. *Seeking spatial justice* (Vol. 16). University of Minnesota Press.

Sunstein, C. R., 2018. #Republic: Divided Democracy in the Age of Social Media. doi.org/10.1111/jcom.12344

Sweeney, L., 2002. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty*, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

Tsou, M. H., Jung, C. T., Allen, C., Yang, J. A., Gawron, J. M., Spitzberg, B. H., Han, S., 2015. Social media analytics and research test-bed (SMART dashboard). In A. Gruzd, J. Jacobson, P.

Twitter, 2022a. Twitter Announces First Quarter 2022 Results. s22.q4cdn.com/826641620/files/doc_financials/2022/q1/Final-Q1'22-earnings-release.pdf.

Twitter 2022b. How to add your location to a Tweet. help.twitter.com/en/using-twitter/tweet-location.

United Nations, Department of Economic and Social Affairs, Population Division, 2019. *World Urbanization Prospects: The 2018 Revision* (ST/ESA/SER.A/420). New York: United Nations.

van Maarseveen, M., Martinez, J., Flacke, J., 2019. *GIS in sustainable urban planning and management: a global perspective* (p. 364). Taylor & Francis.

Voskamp, I. M., Spiller, M., Stremke, S., Bregt, A. K., Vreugdenhil, C., Rijnaarts, H. H., 2018. Space-time information analysis for resource-conscious urban planning and design: A stakeholder based identification of urban metabolism data gaps. *Resources, Conservation and Recycling*, 128, 516-525.

Yeh, A. G., 1999. Urban planning and GIS. *Geographical information systems*, 2(877-888), 1.

Zeleny, M., 1987. Management support systems: towards integrated knowledge management. *Human systems management*, *7*(1), 59-70.

All links last accessed 31/05/2022.