

Availability of open data for spatial public health research

Abstract

Background: Preventive and health-promoting policies can guide (place- and space-specific) factors influencing human health, such as the physical and social environment. Required is data that can lead to a more nuanced decision-making process and identify both existing and future challenges. Along with the rise of new technologies, and thus the multiple opportunities to use and process data, new options have emerged to measure and monitor factors that affect health. Thus, in recent years, several gateways for open data (including governmental and geospatial data) have become available. At present, an increasing number of research institutions as well as (state and private) companies and citizens' initiatives are providing data. However, there is a lack of overviews covering the range of such offerings regarding health. In particular, for geographically differentiated analyses, there are challenges related to data availability at different spatial levels and the growing number of data providers.

Objectives: This paper aims to provide an overview of open data resources available in the context of space and health to date. It also describes the technical and legal conditions for using open data.

Results: An up-to-date summary of results including information on relevant data access and terms of use is provided along with a web visualization. All data is available for further use under an open license.

Keywords: data infrastructure, open data, open government, open geodata, health research

Manuela Peters^{1,2}

Hajo Zeeb^{1,2}

1 Leibniz Institute for
Prevention Research and
Epidemiology – BIPS,
Bremen, Germany

2 Faculty 11 – Human and
Health Sciences, University
of Bremen, Germany

Introduction

There is a close relationship between place, space and health, with each location posing different conditions and challenges. For instance, it is not only socioeconomic conditions, environmental exposures and the quality and character of the natural and built environment that vary, but also access to and accessibility of health-related resources. According to the 'Health in All Policies' approach, building health-promoting living conditions on a local level should be considered in all areas of policy and administration. This has already been initiated with the Ottawa Charter [1]. Simultaneously, the aim is to prevent health impairments through targeted activities within the framework of situational prevention. Specific measures in line with the Public Health Action Cycle require data-based information. Although the Federal Health Reporting (GBE) system, for example, can draw on a range of data sources, the information available varies depending on the geographical unit and the data situation [2]. As a result of digitization, the supply of data, the number of providers and the associated possibilities for data linkage have grown considerably. In particular, *Open Data* (OD) made available by public authorities, private or state-owned companies, research institutions or citizens' initiatives allows novel approaches and options for use. The term *Open Government Data* (OGD) is currently used to describe a growing number of supply channels for admin-

istrative data, along with standards for research and provision, and the offering of selected data through company websites and/or general portals. Compared to other countries, however, Germany is still lagging behind when it comes to exploiting this potential for public health research [3]. One of the main causes is the complexity of the supply market [4]. For instance, only a few local authorities and regional federations run OD portals, which are also only partially integrated into umbrella catalogs or interlinked. Finding data, particularly for specific geographic scales and the subject area discussed in this article, therefore remains challenging. In addition, different technical and regulatory constraints hinder the use of the data. Regarding FAIRification (FAIR is an acronym for findability, accessibility, interoperability and reusability [5]), the problems are very similar to those experienced with research data in general.

Objectives and methodology

This article seeks to provide orientation on currently available data offerings considering space and health on different administrative levels that can be used and disseminated both barrier-free and without any technical and legal restrictions. The selection of relevant explanatory determinants is based on theoretical models. Since the aim of the project is to capture current practice initi-

Table 1: Legal and technical openness

| Legal openness: most common licenses and application contexts | |
|---|---|
| Creative Commons (CC) | Does not solely apply to data and allows works to be copied, distributed, and re-used, according to the terms specified. CC can be used worldwide, it is valid for the duration of copyright protection, and it is often used for geospatial data. Licensors can add permissions to set conditions. Accordingly, the license models differ in versions 1.0 to 4.0, with the latter applying to the international context and being used, for example, by EU data providers. https://creativecommons.org/licenses |
| Open Database License (ODC) | Available in the following versions: <ul style="list-style-type: none"> • Open Data Commons Public Domain Dedication and License (PDDL): all copyrights to data waived • Open Data Commons Attribution License (ODC-BY): sharing and editing permitted; attribution required • Open Data Commons Open Database License (ODC-ODbL): sharing and editing permitted; attribution and sharing under equal terms required http://opendatacommons.org/licenses/odbl/ |
| Data License Germany | Mainly used for OGD, currently available in the following versions: <ul style="list-style-type: none"> • Data License Germany – Attribution – Version 2.0 – requires attribution to the provider according to its terms and the license name • Data License Germany – Zero – Version 2.0 – allows for unrestricted reuse https://www.govdata.de/web/guest/lizenzen |
| Formats for texts, tables and images | <i>.txt, .csv, .htm, .xml, .rdf, .odt, .ods, .rss, .pdf, .doc, .docx, .xls, .xlsx, .rtf, .gif, .jpg, .png, .tif, .geotiff</i> |
| Formats for geodata | <i>.gml, .gpx, .kml, .dxf, .dwg, .shp, .shx, .dbf, .svg</i> |

italic: machine-readable, **bold**: open format, **bold + italic**: open and machine-readable
Own arrangement, technical openness based on [19]

atives, and since hardly any scientific research is available for this context so far, availability of relevant data is examined using a simple web search (Google) and complemented by reviewing references on relevant websites as well as in catalogs and data portals. Based on a snapshot of the current situation, data sources are identified and classified according to theoretical dimensions following a definitional introduction to OD and a description of the conceptual framework. For this purpose, existing freely available datasets (e.g. web visualizations), have been adapted and extended to the new topic. This data is provided for further use under an open license. The information may be relevant for politics, planning and science as well, e.g. as a basis for local activities in the field of prevention and epidemiological studies.

State of research

In Germany, the interdisciplinary discourse on OD (especially on reuse) is still in its infancy stages and mainly driven in the administrative context of OGD, with only a few comprehensive papers [6], [7], [8] or web visualizations [9], [10], [11] available. To date, apart from solely geodata-based reviews [12], [13], [14] and reviews focusing mainly on administrative data sources for public health research [2], [15], [16], [17], there are no existing overviews for the specific context of space-based health research regarding relevant open data sources, their access modes and possible uses.

Open data

Published data is not necessarily open. It is not open, for example, if rights of use are retained or granted on a case-by-case basis only, or if the data is provided in a format preventing further use. By definition, data is open (in technical and legal terms) in particular if it is machine-readable and structured and can be freely used, reused and distributed by anyone using an open interface for any purpose and without any restrictions, discrimination or fees [18]. Efficient data reuse therefore implies that it is available in good quality, as raw data, ideally including metadata, and with clearly stated rights of use. Licensing conditions are addressed within the framework of *legal openness*. Most public data can only be used in accordance with declared conditions, e.g. allowing commercial use, processing and dissemination. For personal data protected by law, subsequent use is not permitted in most cases. Table 1 provides orientation, listing commonly used licenses that allow further processing in the sense of OD, mostly without any restrictions. *Technical openness* relates to data formats ensuring reuse with both proprietary and free software. In recent years, recommendations for the publication of data have been formulated. The OGD principles listed in Table 2 have been considered standard ever since.

Table 2: Open data criteria [56]

| Criteria | Description of requirements |
|-------------------------------|---|
| Completeness | As complete as possible, as permitted by data protection regulations; if possible in raw format, including descriptive metadata |
| Primary sources | From original sources, with information on collection and preparation |
| Timeliness | Publication close to the time the data is created or updated |
| Accessibility | Avoidance/minimization of barriers, e.g. due to technical problems, file formats, missing metadata, organizational barriers (e.g. mandatory registration), legal requirements |
| Machine readability | Use of uniform, machine-readable formats |
| Non-discriminatory | Access without time restrictions or required registrations, no restriction to selected software applications |
| Use of open standards | Use of formats which can be read and processed by more than just specific programs, no need to pay license fees to software publishers |
| Terms of use/licensing | Clear and uniform terms of use, clearly visible license that specifies conditions of use |
| Sustainability | Permanent availability, comprehensible documentation of changes, updates and deletions with version control and archiving |
| Cost of use | No provision against payment |

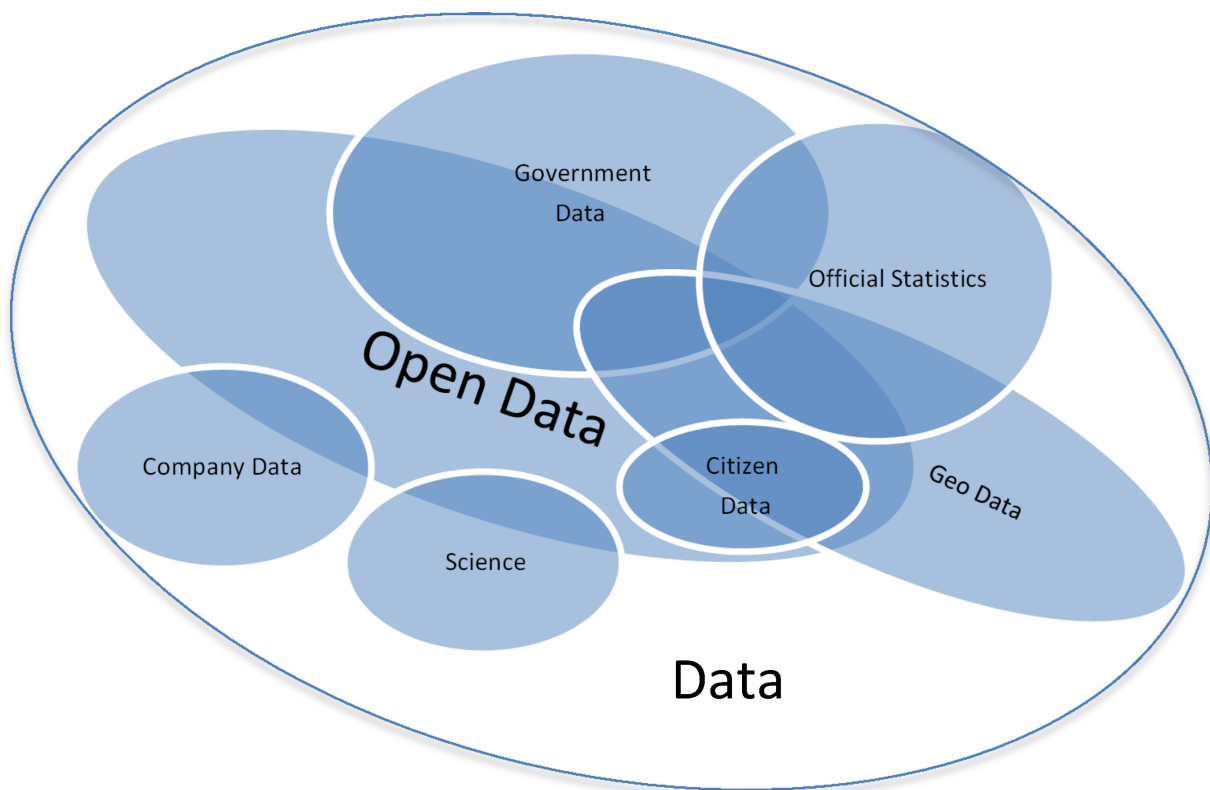


Figure 1: Open data and intersections by data source

Data sources

In addition to OGD, open data that can be used for spatial public health research includes a wide range of open geospatial data as well as data from official statistics, which in some cases are already linked to geospatial

data. Furthermore, companies, research institutions and the civil community offer data in open formats for further use. The sources shown in Figure 1 are described in more detail below.

OGD and subject data

The term OGD refers to its origin in the public sector (authorities, agencies, state departments) [19]. In many cases, such data is available in an aggregated form only, sometimes it is difficult to find or can only be accessed upon request (e.g. based on the Freedom of Information Act (FOIA)). For instance, numerous authorities provide ordered supplies that are usually free of charge but do not fulfill the criteria of OGD by definition. Whereas OGD has so far been offered mainly on a provider-specific basis, in the future it is to be made available proactively, to anyone, for any purpose, without restrictions and free of charge, so that it can be reused and redistributed [20]. The formats provided and the degree of accessibility determine *whether* or as *how* open OGD can be described [19]. To quantify this and as proof of quality, more and more data providers use the cascading 5-star model proposed by Berners-Lee (the second star can only be awarded to those who already fulfill the requirements regarding the first star). Said model is based on the assumption of an open license and topped by data published in the web in a structured, non-proprietary format with a unique URL to link the data [21]. National OGD platforms such as *Open Data Monitor* or *GDI-DE* have been supported by the federal government for several years, combining regional and municipal offerings. The national metadata platform *GovData.de* does not maintain any data itself but refers to OGD from all federal levels. The Federal Ministry of Transport and Digital Infrastructure (BMVI) hosts multiple access points, including *mdm Portal*, *CODE-DE2*, data from the Copernicus Earth Observation Program and *mCLOUD* (for mobility, geospatial and weather data). The latter also serves as a research platform for OD from private providers in the mobility sector. In addition, Germany's National Meteorological Service *Deutscher Wetterdienst* (DWD) currently offers free access to open meteorological and climatological data on its *OD portal*. Some ministries do not make data available themselves but through open data portals.

Open geodata

Open geodata emerge from government, research, the private sector or crowd-sourced initiatives such as *OpenStreetMap* (OSM), and can be made available as thematically prepared geospatial data or basic geodata describing properties and topographies [22]. For public health research, the value of this resource arises primarily through *data linkage* with attribute data [23], [24] and hence the ability to identify ecological correlations or spatial disparities in relation to health-related issues [25]. It is necessary that the data is up to date, exhaustive, quality-checked, provided in uniform and interoperable form (e.g. using the harmonized pan-European metadata standard DCDAT), that it has transparent terms of use and documented metadata, and that it is permanently available [26]. By introducing the Infrastructure for Spatial Information in Europe (INSPIRE), the

legal basis for this has been established. In Germany, this is implemented through the Infrastructure for Spatial Information in Germany (GDI-DE). Central access points to search for open geodata on all federal levels are provided by *GovData.de* and *Geoportal.de*. OD is also offered by the Federal Agency for Cartography and Geodesy (BKG). Map-compatible datasets are provided in regional databases via interoperable download services. However, these are not necessarily OD in every case. Private companies such as ESRI und Google (*GoogleMaps*) also make selected geodata available for private use using free licenses; however, some of them can only be used with proprietary software. Another resource for health-related spatial data, provided by the National Association of Statutory Health Insurance Funds (GKV-Spitzenverband) and the German Hospital Association (DKG), has been in regular operation since January 2020: The web application *krankenhausstandorte.de* allows providers to register geocoded hospital and outpatient facilities (including metadata), which can then be accessed by users (requires registration).

Open data from official statistics

Data from official statistics include geospatial as well as registry, social, environmental and health data. At federal level, the Federal Statistical Office is the leading data provider. It is also possible to obtain official OD from regional databases, local or municipal governments as well as from overarching portals and databases, such as the *INKAR database* of the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). Social welfare statistics and social reporting also contain relevant OD at federal level and, in some cases, at the small-scale level, available at *state or local government gateways*. So far, however, such "atlases of social infrastructure" have mainly been offered in non-machine-readable PDF format.

Open company data

The data portal of *Deutsche Bahn AG* is an example of a private company that offers OD on a national scale via its own gateway. Likewise, the *Open Data Platform ÖPNV* initiative, powered by nine major German transport associations, makes OD freely available on its portal. In addition, more and more regional and local transportation services are offering open data on their corporate websites (e.g. *NVBW Open Data*).

Open data through Citizen Science (CS)

In addition to government activities, volunteers collect data in crowd-sourced projects and make it available for free (further) use under open licenses. Well-known initiatives include the projects of the OK-Lab Stuttgart with the <https://sensor.community/de/> and *OpenSenseMap.org* platforms as well as some projects of the *Codefor-Lab* (<https://www.codefor.de/projekte/>). The *offenedaten.de*

portal, which has, however, been closed in the meantime and on which Germany-wide OD was mapped interactively for the first time on the basis of web-crawler matches, still serves as a prototype for such applications today [13]. Similar concepts are pursued by the *OpenDataAtlas* [9] and *Opendata-City-Census*. The potential of these resources becomes particularly obvious when it comes to aspects such as copyright, terms of use, costs and availability (e.g. of OD missing from the official side or not recorded or offered in a uniform manner).

Open research data (Open Science)

To some extent, open research data may also be re-used as well as linked to spatial data. A beneficial development in this context is the growing commitment to *FAIR Data Principles* [5]. However, data meeting these principles does not necessarily have to be open.

(Health) surveys are another potential source of data. To perform regionalization of national surveys, the samples need to be very large, so no usable area-wide, small-scale data is available for many aspects of interest. In some cases, the data is restricted to scientific purposes and therefore available as Scientific Use Files (SUF) only. An international register with detailed information on more than 2,000 research data repositories with open data for re-use, under defined conditions, is provided by the Registry of *Research Data Repositories* platform (*re3data.org*).

Conceptual framework

Healthy living conditions are a multidimensional concept with complex interactions. Whereas the main focus was on individual [27], [28], [29] and environmental [30], [31], [32] risk factors in the past, today protective determinants in the sense of salutogenesis, such as social structures, availability of environmental resources (e.g. green spaces) and potential factors that may affect physical activity are increasingly taken into account [33], [34]. The holistic human ecology paradigm [35], which identifies influencing factors from social, ecological and economic systems, can be used to organize the data. The dimensions are in line with the aspects that the World Health Organization (WHO) has identified as essential for human well-being based on a review [36]. In addition to individual (lifestyle) factors, these include structural conditions of the living environment, such as (1) sociodemographic and socioeconomic factors, (2) factors of access and reachability, (3) the physical neighborhood, and (4) climate and environmental factors. This theoretical framework serves as the foundation for selecting and organizing the data as presented in the following data research results.

Results: relevant dimensions and data sources

Although full datasets are available in the field of *environment and health* at national and municipal levels, e.g. from environmental reporting (UBE) and the federal/state indicators of health monitoring, they do not refer to each other [37]. In addition, the data is mainly available in non-open reporting formats. In accordance with the dimensions previously defined, however, the specific data resources focused on below and listed in Attachment 1, Table 3 and the corresponding web visualization (<https://sciencemap.github.io/Open-Data-for-Health>) can be considered.

Social-environmental indicators: social condition, sociodemographic and socioeconomic factors

Different levels of exposure and thus health effects are, among other factors, determined by the population's socioeconomic and sociodemographic composition [38], [39], [40], with social inequality increasing environmental injustice [41]. For epidemiology, the factors education, occupational status and income are defined as the most important characteristics within this dimension [17], [42]. Furthermore, the economic environment, the financial situation of the public sector, and the level of collective wealth reflect the social capital and the capacity of a region to provide services of public interest. A valuable synopsis of data resources along this dimension, including official statistics, welfare statistics and the GBE, has been elaborated by Bardehle et al. [17]. Information on socio-spatial differences in health can be obtained by linking this data with geodata. Moreover, aggregated indicator systems for measuring socioeconomic deprivation have been established for Germany in the past [43]. Of these, the data provided by the German Index of Socioeconomic Deprivation (GISD) [44], for example, are publicly available and presented in machine-readable form; the license, however, allows no further processing.

Well-being can also be affected by personal feelings of vulnerability (e.g. due to crime and traffic safety). Although the Federal Criminal Police Office (BKA) stopped providing its federal-level police crime statistics (PKS) exclusively as PDF files in 2014, small-scale data in machine-readable form is often missing or can only be accessed upon request.

Availability and accessibility

Access to (care) services relates, among other things, to both general availability and access in the sense of reachability. Social conditions (participation, inclusion) and thus existing social infrastructures are of great importance to health due to their salutogenic protective function [45].

Life expectancy at birth can be interpreted as an indicator of the general living and healthcare conditions in a given region. While family physicians represent a key institution in outpatient healthcare, recent years have witnessed an increasing need for certain healthcare specialists [46]. In addition, both availability and accessibility of facilities for physical activity and community socialization play a central role [47]. Apart from OD derived from official statistics, geodata has become a valuable resource for mapping healthcare supply realities at different spatial levels.

Physical environment

The physical environment has a major impact on health, particularly on a small-scale level. This refers to the mental, physical and social components [48]. The land use mix, i.e. the proportion of open spaces, traffic areas, buildings and important supply infrastructure, is a key factor in this context. Regional and local cadastral offices mostly provide information on public green spaces as OD. Moreover, openly accessible indicators such as land coverage, building density and soil sealing can reflect these aspects, while providing insights into the ecological situation and climate-relevant stresses. For instance, a high degree of sealing possibly harbors a local potential for overheating and thus the risk of heat stress. Last but not least, extensive industrial areas point towards an increased production of exhaust gases [49]. In contrast, low-density areas with less soil sealing are often characterized by a high proportion of green space, which serves as a filter for pollutants and a source of oxygen, reducing air temperature and humidity. When it comes to open, green and recreational spaces, there are several OD resources (Attachment 1, Table 3).

Climate and environment

Environmental hazards are physical, chemical and biological factors affecting individuals externally. Of these, air pollution is considered to contribute to diseases with the greatest attributable burden of disease [50]. Commonly established indicators for measuring the exposure include the emission of the substances sulfur dioxide (SO₂), nitrogen oxides (NO_x), ammonia (NH₃) and particulate matter. Based on the Federal Emission Control Act (BImSchG), measuring stations are maintained to determine and monitor air pollution levels, especially in regions with high traffic volumes (including large cities and metropolitan areas). Information on the spatial dispersion of such pollutant concentrations is partly provided through geostatistical interpolations.

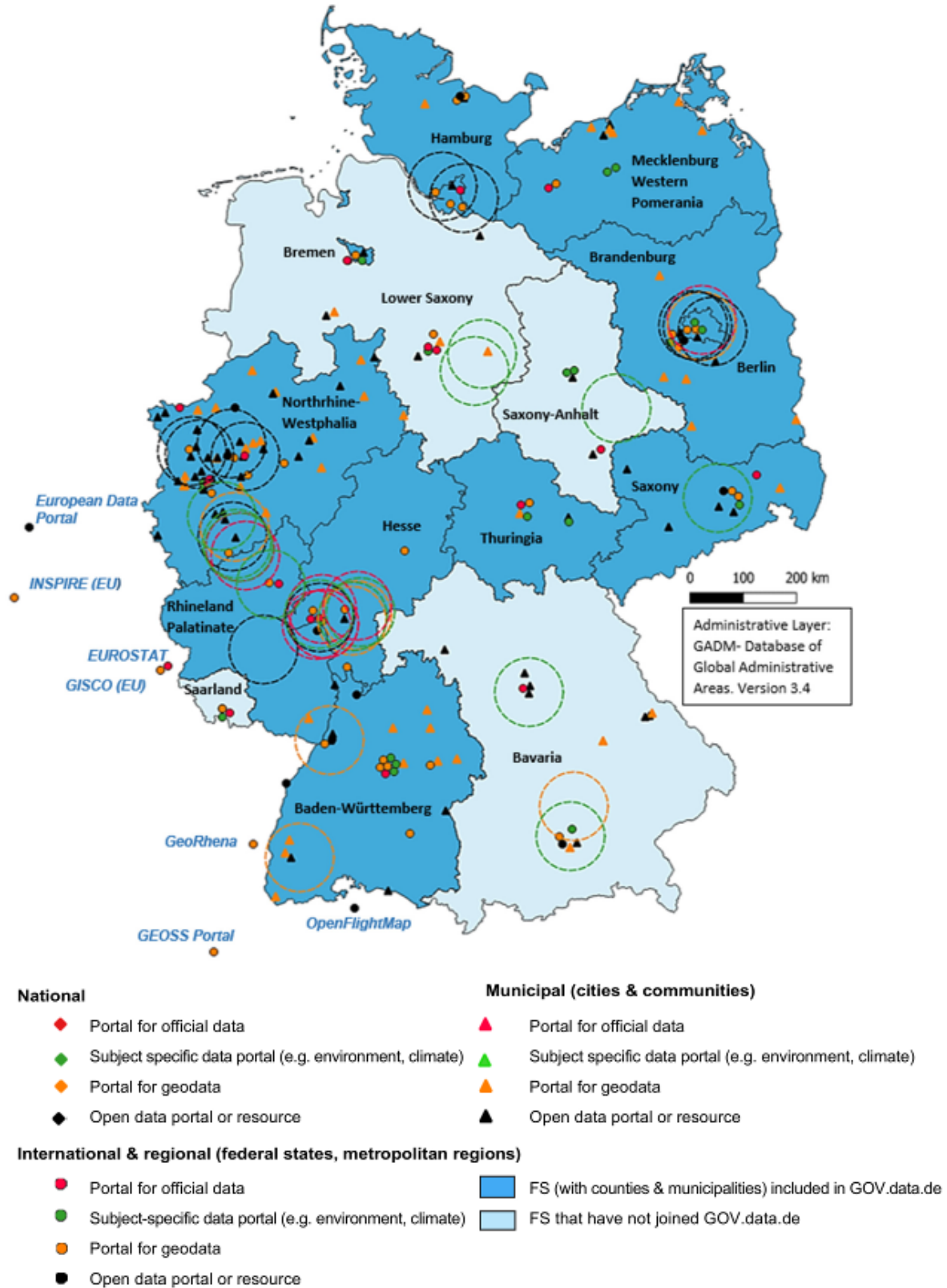
However, with proximity to traffic being a relevant criterion, total annual emissions (in metric tons) in pre-defined administrative units are of restricted explanatory value. Thus, there is limited applicability for deriving area-wide, small-scale exposures (e.g. daily concentrations for fine particulate matter). This also applies to other data collected only at specific measuring points, such as

meteorological data of the DWD. Direct environmental influences also appear as health consequences due to climate or environmental noise. Noise maps for road traffic are usually available for large cities with more than 100,000 inhabitants, in some cases as comprehensive model calculations and with data from noise appraisals. For smaller cities and communities, however, the situation is more varied. Likewise, climate data (and climate function maps) are available in some cases, especially for small-scale areas. Soil and water quality data (factors affecting health indirectly) is largely available at most scales.

Spatial level and data availability

Although there is political and public interest in implementing health-promoting policies, especially on a municipal level, problems arise especially when it comes to small-scale analyses. For data-protection reasons, raw data is not provided for the smallest geographical units but processed or aggregated for specific purposes instead. Moreover, different administrative boundaries and aggregations make it difficult to compare and link data with attribute data. Besides said limited availability and reduced ability for regionalization, data usage is mainly hampered by federal structures. For instance, central portals such as *Gov.Data* can only host municipal datasets if the associated federal states (FS) support the portal. To date, however, only 12 FS have joined the agreement. Instead, some federal states have launched their own data portals. As a result, the amount of data, access to it and the legal framework are heterogeneous and, in some cases, do not come with any legal regulations based on the Freedom of Information Act (IFG). Even the municipal GBE is focused on a state- or region-specific perspective and does not exhibit a common basic standard.

Despite the slight increase in the number of municipalities publishing OD (especially in North Rhine-Westphalia, Hamburg and Berlin), the provision is currently still very selective or, in some regions, almost non-existent (Figure 2). At different spatial scales, some parties provide extensive OD, while others offer thematically non-comparable OD or no OD at all (especially in low-finance local areas). In 2021, out of more than 11,000 municipalities in Germany, only about 146 made any data available at all, some of them with less than 10 datasets [51]. Accordingly, the number is further reduced for the context discussed here. Attachment 1, Table 4 provides an overview of data resources, data types and licenses by different administrative level, which could have been identified based on this current research.



Note: In each case, the geocoded location of the data provider is mapped. A number of coordinates have been shifted to avoid overlapping in the visualization. The icons for the national portals (rings) do not represent geographic coverage but have been selected for illustrative purposes only. Further information, including references to the resources mentioned, as well as an overview of small-scale data portals (city level) can be found online in the web visualization "Open Data for Health", which is part of the publication: <https://sciencemap.github.io/Open-Data-for-Health/>.

Figure 2: Availability of open data by spatial scale and data source (up to April 2021)

Discussion

Availability, interdisciplinary use and linkage of data are essential for the development and monitoring of health strategies and are in line with the understanding of *health in all policies*. Therefore, public health research seeks to make 'data for action' more visible, accessible and disseminable [52] to identify the need for political action. In the light of ongoing digitization and the resulting increase in data quantity, quality and use, extensive opportunities have opened up for analyzing associations between space and health, ranging from regional to international levels.

In this context, open access to data is a necessary requirement. The data resources highlighted above demonstrate that much data, in particular open geodata, is available in principle, but also that information on how to access it is crucial. Furthermore, in many cases it remains unclear what data is actually available as OD. It is difficult to determine the current supply status of OD, as there is no central or official body to collect this information.

Due to some limitations, the increasing number of different portals provides, at best, an entry point for research. Freedom of information and data-protection laws, which are either specific to each portal or lacking, represent major complicating factors.

Alternate access via search-engine queries, however, first requires indexing of the data, e.g. based on technical openness and accompanied by metadata [12]. The technical, legal and content-related heterogeneity of the offerings discussed above as well as regulatory conditions still hinder the systematic use and thus full exploitation of the potential offered by these resources. For instance, a variety of data is accessible in principle but cannot be retrieved in machine-readable, open form.

Furthermore, both quantity and quality of the available data vary. In particular, small-scale and comprehensive availability is patchy and highly heterogeneous. Despite some initiatives such as INSPIRE, finding and accessing geospatial data also remains somewhat challenging. Since health-related geographic data is relatively scarce [53], surveillance by the Robert Koch Institute, for example, hardly takes any advantage of INSPIRE so far [25]. Data openness also covers aspects of participation and continued (re)usage. Although administrative data providers are encouraged to share raw data in accordance with the E-Government Act, the decision on how to use the data still remains with the authorities themselves. In addition, the absence of an explicit license prevents further use in many cases. The Geodata Access Act (GeoZG) also calls for no-cost data sharing but imposes restricted reuse that often incurs a fee. Thus, there is still a mixed landscape in terms of fees and (licensing) policies for geodata, which makes it difficult to merge data in a comprehensive manner. On the one hand, Open Citizen projects such as OSM are a valuable source in this context, not just in terms of data volume. On the other hand, this data exhibits certain disadvantages in terms of completeness, quality, topicality and validity. In contrast,

official geodata is often available for administrative levels but not applicable to scientific analyses. As an example, small-scale epidemiology data is frequently provided on a postal code level, which overlaps inconsistently with administrative units. Other challenges of administrative geodata affecting spatial analyses and calling for alternative approaches are territorial reforms, data aggregations, scaling and zoning effects. In some cases, even a change of indicators may become necessary for area-wide or cross-level analyses for availability reasons.

Outlook

For various reasons, it is currently not possible to give an exhaustive overview of OD offerings. Given that such overviews are only able to represent a snapshot in a dynamically changing context, it would be useful to expand them on an ongoing basis. A desirable feature might be portals/overviews using software to import relevant, ideally standardized, data catalogs from decentralized resources. Metadata (including technical and legal conditions) is a key basis and requires the use of (meta)data catalogs as well as unique identifiers for data records supplied by all data providers.

The provision of OD is a joint process involving several micro and macro levels (state, federal states, counties, municipalities, institutions). It also implies that, besides urgently needed harmonization of legal aspects, general processes must be redesigned starting already in the context of (official) data collection and research to initially give all data the option of openness and machine readability. In particular, the institutional and administrative bodies responsible for quality control, data management and data protection should be integrated into such strategies and processes at an early stage.

Regarding the further use of research data in the course of FAIRification, discipline-specific efforts to achieve common standards, e.g. ontologies and vocabularies, would have to be pushed forward to optimize the findability and indexing of data and metadata sets in both thematic databases and general search engines. To improve the performance of the latter, Google, Yahoo, Yandex and Bing have already developed the *schema.org* metadata standard, which helps data-producing and data-providing parties structure data according to uniform vocabulary. The increased integration of Linked Data standards with the delivery process can also significantly improve data findability, accessibility and comparability in the future.

To ensure the potential further use of data, (harmonized) legal foundations are required on the level of the federal states and municipalities. Orientation for OGD on the municipal level is currently provided by the model data catalog, which lists types and sources of OGD [6].

Concerning OGD, the draft law dated February 2021 on the *Second Open Data Act and Data Utilization Act* (Section 12a EGovG) [54] could have been an important milestone. Although its passing (of June 24, 2021) oblig-

ates authorities of the intermediate federal administration to offer unprocessed, machine-readable data to the public, the claim upon this provision (also for non-protected, personal data) is still restricted. Data from private companies (including public transport) and publicly funded research data, which are newly included in this law, are also partly affected by extensive, non-OD-conform restrictions. For instance, the offering parties may charge fees for their data and decide on licensing (and thus on further use). Consequently, one overarching future goal must be to create general incentives for OD – not least in terms of financial and human resources. Within the public health research landscape, the growing potential of data sharing, despite the (initial) costs of FAIRification and the opening of data in line with OD principles, should be acknowledged, actively requested and used.

Abbreviations

- BBSR: Bundesinstitut für Bau-, Stadt- und Raumforschung (Federal Institute for Building, Urban and Spatial Research)
- BImSchG: Bundesimmissionsschutzgesetz (Federal Emission Control Act)
- BKA: Bundeskriminalamt (Federal Criminal Police Office)
- BKG: Bundesamt für Kartographie und Geodäsie (Federal Agency for Cartography and Geodesy)
- BMVI: Bundesministerium für Verkehr und digitale Infrastruktur (Federal Ministry of Transport and Digital Infrastructure)
- CC: Creative Commons
- DKG: Deutsche Krankenhausgesellschaft (German Hospital Association)
- DWD: Deutscher Wetterdienst (Germany's National Meteorological Service)
- FOIA: Freedom of Information Act
- FS: Federal States
- GBE: Gesundheitsberichterstattung (Federal Health Reporting)
- GDI-DE: Infrastructure for Spatial Information in Germany
- GeoZG: Geodatenzugangsgesetz (Geodata Access Act)
- GISD: German Index of Socioeconomic Deprivation
- INKAR: Indikatoren und Karten zur Raum- und Stadtentwicklung (Indicators and maps for spatial and urban development)
- INSPIRE: Infrastructure for Spatial Information in Europe
- OD: open data
- ODbL: Open Database License
- ODC: Open Data Commons
- ÖPNV: Öffentlicher Personennahverkehr (public transport)
- OGD: open government data
- OSM: OpenStreetMap
- PDDL: Public Domain Dedication and License
- PKS: polizeiliche Kriminalitätsstatistik (federal-level police crime statistics)
- SUF: Scientific Use Files
- UBE: Umweltberichterstattung (environmental reporting)
- WHO: World Health Organization

Links to the resources mentioned within the text

- CODE-DE2: <https://code-de.org/en/>
- Data portal of Deutsche Bahn AG: <https://data.deutschebahn.com/>
- ESRI Open Data: <https://opendata-esri-de.opendata.arcgis.com/>
- Google Maps: <https://about.google/brand-resource-center/products-and-services/geo-guidelines/>
- Hospital Locations: <https://krankenhausstandorte.de>
- INKAR: <https://www.inkar.de/>
- mCLOUD: <https://www.mcloud.de/>
- mdm Portal: <https://www.mdm-portal.de/>
- National social welfare statistics: <https://www.statistikportal.de/de/sbe>
- NVBW (Local transport company Baden-Württemberg) Open Data: <https://www.nvbw.de/open-data>
- OD portal of the German Weather Service: https://www.dwd.de/EN/ourservices/_functions/search/search_Formular.html
- OpenDataAtlas: <http://opendata.tursics.de/>
- Opendata-City-Census: <http://de-city.census.okfn.org>
- Open Data for Health Map: <https://sciencemap.github.io/Open-Data-for-Health/>
- Open Data Platform ÖPNV: <https://www.opendata-oePNV.de>
- OpenSenseMap: <https://OpenSenseMap.org>
- OpenStreetMaps: <https://www.openstreetmap.org>
- Regional social welfare statistics (examples): <https://statistikportal.thueringen.de/thonsa/SSDstart.php> and <http://www.gsi-berlin.info/>
- Registry of Research Data Repositories: <http://re3data.org>
- Sensor Community: <https://sensor.community/en/>

Data

Data for this article are available from the Dryad Repository: <https://doi.org/10.5061/dryad.djh9w0w1f> [55].

Notes

Acknowledgments

The authors would like to thank NFDI4Health for the inspiration and knowledge sharing. This was of great value for the whole work process.

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from <https://doi.org/10.3205/000303>

- Attachment1_000303.pdf (167 KB)
Availability of data in the context of space and health, Overview of data access

References

- Eriksson M, Lindström B. A salutogenic interpretation of the Ottawa Charter. *Health Promot Int.* 2008 Jun;23(2):190-9. DOI: 10.1093/heapro/dan014
- Trojan A. Integrierte Gesundheitsberichterstattung auf Landes- und kommunaler Ebene: Initiativen und Ansätze der letzten 20 Jahre [Integrated health reporting at the communal and federal state level-policy initiatives and approaches of the last 20 years]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2020 Sep;63(9):1084-93. DOI: 10.1007/s00103-020-03196-7
- Burgun A, Bernal-Delgado E, Kuchinke W, van Staa T, Cunningham J, Lettieri E, Mazzali C, Oksen D, Estupiñan F, Barone A, Chène G. Health Data for Public Health: Towards New Ways of Combining Data Sources to Support Research Efforts in Europe. *Yearb Med Inform.* 2017 Aug;26(1):235-40. DOI: 10.15265/IY-2017-034
- Kubicek H, Lippa B. Open Data: Appelle reichen nicht. *Kommune21.* 2015 Mar;15(5):12-3. Available from: https://www.kommune21.de/meldung_21279.html
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016 Mar;3:160018. DOI: 10.1038/sdata.2016.18
- Krabina B, Wiedemann M. Open Data in Kommunen. Welche Daten stellen Kommunen als Open Data zur Verfügung? Ein Musterdatenkatalog für Nordrhein-Westfalen. Gütersloh: Bertelsmann Stiftung; 2019. (Analysen und Konzepte – LebensWerte Kommune; 2/2019). Available from: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Smart_Country/AK_2_2019_05_08_ONLINE.pdf
- Begleitforschung Smart Service Welt II, editor. Open Public Data in Deutschland. Rahmenbedingungen und Potenziale der Bereitstellung und Nutzung von Daten des öffentlichen Sektors. Berlin: Institut für Innovation und Technik; 2020. Available from: https://www.iit-berlin.de/iit-docs/e76a033d9c274cf1a65efab271552894_SSW_Open_Public_Data_in_Deutschland2.pdf
- Knetsch G, Fock J. Open Government Data – Zugang zu Umwelt- und Gesundheitsdaten. *UMID.* 2013;(3):20-4. Available from: https://www.umweltbundesamt.de/sites/default/files/medien/360/publikationen/open_government_data_s_20_24n.pdf
- Tursics T. Open Data Atlas. [last accessed 2021 Mar 29]. Available from: <https://opendata.tursics.de/>
- OpenGeoEdu. OpenDataPortal. [last accessed 2021 Mar 29]. Available from: <https://portal.opengeoedu.de/>
- Open Knowledge Foundation Deutschland e.V. Offene Daten. [last accessed 2021 Mar 29]. Available from: <https://okfn.de/en/projekte/offenedaten/>
- Paderta D. Open Data – Raumbezogene Daten. Köln: Gesis; 2012. Available from: https://www.ssoar.info/ssoar/bitstream/handle/document/36474/ssoar-2012-paderta-Open_Data_-_Raumbezogene_Daten.pdf
- Hinz M, Bill R. Ein zentraler Einstiegspunkt für die Suche nach offenen Geodaten im deutschsprachigen Raum. *AGIT J Angew Geoinf.* 2018;(4):298-307. DOI: 10.14627/537647038
- Kalasek R, Weninger K. Open GeoData. *Oeff Sect.* 2015 Apr;41(1):17-28. DOI: 10.34749/oes.2015.1274
- Fehr A, Lange C, Fuchs J, Neuhauser H, Schmitz R. Gesundheitsmonitoring und Gesundheitsindikatoren in Europa. *J Health Monit.* 2017 Mar;2(1):3-23. DOI: 10.17886/RKI-GBE-2017-004.2
- Bardehle D, Anuss R, Hermann S, Ziese T, Böhm K. Der neue Länderindikatorenset für die Gesundheitsberichterstattung [The new indicator set for health reporting activities in the German States]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2004 Aug;47(8):762-70. DOI: 10.1007/s00103-004-0872-x
- Bardehle D, Blettner M, Laaser U. Gesundheits- und soziodemographische (sozialepidemiologische) Indikatoren in der Gesundheits- und Sozialberichterstattung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2001 Apr;44(4):382-93. DOI: 10.1007/s0010300050456
- Open Knowledge Foundation Deutschland. Das ABC der Offenheit. Berlin: OKFN/Wikimedia; 2019. Available from: https://upload.wikimedia.org/wikipedia/commons/a/a9/ABC_der_Offenheit_-_Broschüre_%282019%29.pdf
- von Lucke J, Geiger C. Open Government Data – Frei verfügbare Daten des öffentlichen Sektors. In: von Lucke J, editor. Entdeckung, Erkundung und Entwicklung 2.0: Open Government, Open Government Data und Open Budget 2.0. Friedrichshafen: Band 1 der Schriftenreihe des Deutsche Telekom Institute for Connected Cities, TICC der Zeppelin Universität; 2012. p. 105-56.
- Internet und Gesellschaft Collaboratory. Offene Staatskunst – Bessere Politik durch „Open Government“? 1. Auflage; 2010. Available from: <http://d-nb.info/1120152208/34>
- Janowicz K, Hitzler P, Adams B, Kolas, D, Vardemann C. Five stars of linked data vocabulary use. *Semantic Web.* 2014;5(3):1-3.
- Meinel G. Grundlagen der Geo-Informationssysteme. *Raumforsch Raumordn.* 2016 Dec;74(6):593-4. DOI: 10.1007/s13147-016-0418-3
- Fletcher-Lartey SM, Caprarelli G. Application of GIS technology in public health: successes and challenges. *Parasitology.* 2016 Apr;143(4):401-15. DOI: 10.1017/S0031182015001869
- Nykiforuk CI, Flaman LM. Geographic information systems (GIS) for Health Promotion and Public Health: a review. *Health Promot Pract.* 2011 Jan;12(1):63-73. DOI: 10.1177/1524839909334624

25. Thißen M, Niemann H, Varnaccia G, Rommel A, Teti A, Butschalowsky H, Manz K, Finger JD, Kroll LE, Ziese T. Welches Potenzial haben Geoinformationssysteme für das bevölkerungsweite Gesundheitsmonitoring in Deutschland? Perspektiven und Herausforderungen für das Gesundheitsmonitoring am Robert Koch-Institut [What potential do geographic information systems have for population-wide health monitoring in Germany? Perspectives and challenges for the health monitoring of the Robert Koch Institute]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2017 Dec;60(12):1440-52. DOI: 10.1007/s00103-017-2652-4
26. Kany C, Klein U, Osterhold M, Riecken J, Sandmann S, Schaffert M, Schön B, Seuß R. Wert von Geoinformation. *Z Geodäsie Geoinf Landmanag*. 2018;143(6):390-7. DOI: 10.12902/zfv-0238-2018
27. Kaplan MS, Newsom JT, McFarland BH, Lu L. Demographic and psychosocial correlates of physical activity in late life. *Am J Prev Med*. 2001 Nov;21(4):306-12. DOI: 10.1016/s0749-3797(01)00364-6
28. Lim K, Taylor L. Factors associated with physical activity among older people – a population-based study. *Prev Med*. 2005 Jan;40(1):33-40. DOI: 10.1016/j.ypmed.2004.04.046
29. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. *BMC Public Health*. 2013 May;13:449. DOI: 10.1186/1471-2458-13-449
30. Stokols D, Grzywacz JG, McMahan S, Phillips K. Increasing the health promotive capacity of human environments. *Am J Health Promot*. 2003 Sep-Oct;18(1):4-13. DOI: 10.4278/0890-1171-18.1.4
31. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006 Apr;27:297-322. DOI: 10.1146/annurev.publhealth.27.021405.102100
32. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. *Am Psychol*. 1992 Jan;47(1):6-22. DOI: 10.1037//0003-066x.47.1.6
33. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens B, Bachmann W. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *J Am Plann Assoc*. 2006;72(1):75-87. DOI: 10.1080/01944360608976725
34. Giles-Corti B, Vernez-Moudon A, Reis R, Turrell G, Dannenberg AL, Badland H, Foster S, Lowe M, Sallis JF, Stevenson M, Owen N. City planning and population health: a global challenge. *Lancet*. 2016 Dec;388(10062):2912-24. DOI: 10.1016/S0140-6736(16)30066-6
35. Barton H, Grant M. A health map for the local human habitat. *J R Soc Promot Health*. 2006 Nov;126(6):252-3. DOI: 10.1177/1466424006070466
36. Neira M, Prüss-Ustün A. Preventing disease through healthy environments: A global assessment of the environmental burden of disease. *Toxicol Lett*. 2016 Oct;259:S1. DOI: 10.1016/j.toxlet.2016.07.028
37. Tobollik M, Kabel C, Mekel O, Hornberg C, Plaß D. Übersicht zu Indikatoren im Kontext Umwelt und Gesundheit [Overview of indicators in the context of environment and health]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2018 Jun;61(6):710-9. DOI: 10.1007/s00103-018-2743-x
38. Hofmeister C, Maier W, Mielck A, Stahl L, Breckenkamp J, Razum O. Regionale Deprivation in Deutschland: Bundesweite Analyse des Zusammenhangs mit Mortalität unter Verwendung des ‚German Index of Multiple Deprivation (GIMD)‘ [Regional Deprivation in Germany: Nation-wide Analysis of its Association with Mortality Using the German Index of Multiple Deprivation (GIMD)]. *Gesundheitswesen*. 2016 Jan;78(1):42-8. DOI: 10.1055/s-0034-1390421
39. Flacke J, Schüle SA, Köckler H, Bolte G. Mapping Environmental Inequalities Relevant for Health for Informing Urban Planning Interventions – A Case Study in the City of Dortmund, Germany. *Int J Environ Res Public Health*. 2016 Jul;13(7):711. DOI: 10.3390/ijerph13070711
40. Briggs D. Environmental pollution and the global burden of disease. *Br Med Bull*. 2003 Dec;68(1):1-24. DOI: 10.1093/bmb/ldg019
41. Bolte G, Bunge C, Hornberg C, Köckler H. Umweltgerechtigkeit als Ansatz zur Verringerung sozialer Ungleichheiten bei Umwelt und Gesundheit [Environmental justice as an approach to tackle environmental health inequalities]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2018 Jun;61(6):674-83. DOI: 10.1007/s00103-018-2739-6
42. Ahrens W, Bellach BM, Jöckle KH. Messung soziodemographischer Merkmale in der Epidemiologie. München: MMV Medizin Verlag; 1998. (RKI-Schriften; 1/98).
43. Maier W. Indizes Multipler Deprivation zur Analyse regionaler Gesundheitsunterschiede in Deutschland: Erfahrungen aus Epidemiologie und Versorgungsforschung [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany: Experiences from epidemiology and healthcare research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2017 Dec;60(12):1403-12. DOI: 10.1007/s00103-017-2646-2
44. Kroll LE, Schumann M, Hoebel J, Lampert T. Regionale Unterschiede in der Gesundheit – Entwicklung eines sozioökonomischen Deprivationsindex für Deutschland. *Journal of Health Monitoring*. 2017;2(2):103-20. DOI: 10.17886/RKI-GBE-2017-035
45. Gruenewald TL, Karlamangla AS, Greendale GA, Singer BH, Seeman TE. Feelings of Usefulness to Others, Disability, and Mortality in Older Adults: the MacArthur Study of Successful Aging. *J Gerontol B Psychol Sci Soc Sci*. 2007 Jan;62(1):P28-37. DOI: 10.1093/geronb/62.1.p28
46. Walter U, Schneider N, Bisson S. Krankheitslast und Gesundheit im Alter. Herausforderungen für die Prävention und gesundheitliche Versorgung [Morbidity and health in old age. A challenge for prevention and health care]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2006 Jun;49(6):537-46. DOI: 10.1007/s00103-006-1267-y
47. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW; Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012 Jul;380(9838):258-71. DOI: 10.1016/S0140-6736(12)60735-1
48. Claßen T. Urbane Grün und Freiräume-Ressourcen einer gesundheitsförderlichen Stadtentwicklung. In: Baumgart S, Köckler H, Ritzinger A, Rüdiger A, editors. *Planung für gesundheitsfördernde Städte*. Hannover: Verlag der ARL; 2018. p. 297-313.
49. Hartig T, Evans GW, Jamner LD, Davis DS, Gärling T. Tracking restoration in natural and urban field settings. *J Environ Psychol*. 2003 Jun;23(2):109-23. DOI: 10.1016/S0272-4944(02)00109-3

50. United States Environmental Protection Agency. Integrated Science Assessment for Oxides for Nitrogen – Health Criteria. Durham, NC: EPA; 2016. Available from: http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=526855
51. Bertelsmann Stiftung. Musterdatenkatalog schafft erstmals Übersicht über offene Daten in Deutschlands Kommunen. 2021 Mar 3. Available from: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Presse/210303_Pressemitteilung_Musterdatenkatalog_schafft_erstmals_Uebersicht_ueber_offene_Daten_in_Deutschlands_Kommunen.pdf
52. Zens M, Shajanian Zarneh Y, Dolle J, De Bock F. Digital Public Health – Hebel für Capacity Building in der kommunalen Gesundheitsförderung: Ausgangslage, Entwicklungsfragen, TEAviisari als modellhafte Implementierung [Digital public health-leverage for community capacity building in health promotion: Current situation, developmental issues and TEAviisari as a model implementation]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2020 Jun;63(6):729-40. DOI: 10.1007/s00103-020-03148-1
53. Seiler M. INSPIRE-Daten im Gesundheitswesen – Status und Potenzial. In: Strobl J, Blaschke T, Griesebner G, Zagel B, editors. Angewandte Geoinformatik. Berlin/Offenbach: Herbert Wichmann Verlag; 2014. p. 303-8.
54. Bundesministerium der Justiz. Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz - EGovG) – § 12a Offene Daten des Bundes, Verordnungsermächtigung. [last accessed 2022 Mar 1]. Available from: https://www.gesetze-im-internet.de/egovg/_12a.html
55. Peters M, Zeeb H. Data from: Availability of open data for spatial public health research. Dryad Digital Repository. 2022. DOI: 10.5061/dryad.djh9w0w1f
56. Sunlight Foundation. Ten Principles for Opening Up Government Information. 2010 Aug 11. Available from: <https://sunlightfoundation.com/wp-content/uploads/sites/2/2016/11/Ten-Principles-for-Opening-Up-Government-Data.pdf>

Corresponding author:

Manuela Peters

Leibniz Institute for Prevention Research and Epidemiology – BIPS, Achterstraße 30, 28215 Bremen, Germany, Phone: +49 421 218-56924
mpeters@leibniz-bips.de

Please cite as

Peters M, Zeeb H. Availability of open data for spatial public health research. *GMS Ger Med Sci.* 2022;20:Doc01.

DOI: 10.3205/000303, URN: urn:nbn:de:0183-0003030

This article is freely available from

<https://doi.org/10.3205/000303>

Received: 2021-04-19

Revised: 2021-11-11

Published: 2022-03-04

Copyright

©2022 Peters et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Verfügbarkeit offener Daten für die raumbezogene Public-Health-Forschung

Zusammenfassung

Hintergrund: Präventive und gesundheitsförderliche politische Maßnahmen können (raumspezifische) Einflussfaktoren auf die menschliche Gesundheit, wie z.B. die der physischen und sozialen Umwelt, steuern. Grundlage sind Daten, die zu einer nuancierten Entscheidungsfindung führen und sowohl existierende als auch zukünftige Herausforderungen identifizieren können. Mit der Einführung neuer Technologien und damit der vielfältigen Möglichkeiten zur Nutzung und Weiterverarbeitung von Daten haben sich neue Möglichkeiten ergeben, Faktoren, die sich auf die Gesundheit auswirken, zu messen und zu monitorieren. So sind in den letzten Jahren Portale für offene Daten (inklusive Verwaltungs- und Geodaten) entstanden. Immer mehr Forschungseinrichtungen sowie (staatliche und private) Unternehmen und Bürgerinitiativen stellen zudem Daten zur Verfügung.

Ziel der Arbeit/Methoden: Bislang fehlt es allerdings an Übersichten, die die aktuelle Bandbreite dieser Angebote im Hinblick auf den Gesundheitskontext erfassen. Insbesondere für räumlich differenzierte Betrachtungen ergeben sich Herausforderungen bezüglich der Datenverfügbarkeit auf den verschiedenen räumlichen Ebenen und der zunehmend unübersichtlich werdenden, heterogenen Palette der Anbieter. Dieser Beitrag soll, basierend auf Web- und Datenbankrecherchen, einen Überblick geben, wo bislang offene Daten im Kontext von Raum und Gesundheit angeboten werden und unter welchen technischen und rechtlichen Bedingungen diese verwendet werden können.

Ergebnisse: Eine aktuelle Ergebnisübersicht zeigt die stark heterogene Verfügbarkeit relevanter Datenzugänge und unterschiedliche Nutzungsbedingungen relevanter, raumbezogener Public-Health-Daten auf. Die deskriptive Darstellung wird durch eine detaillierte Webvisualisierung unterstützt und steht inklusive aller zugrundeliegenden Daten barrierefrei und unter einer offenen Lizenz allen Interessierten zur Weiternutzung zur Verfügung.

Schlüsselwörter: Dateninfrastrukturen, offene Daten, offene Verwaltungsdaten, offene Geodaten, Gesundheitsforschung

Manuela Peters^{1,2}
Hajo Zeeb^{1,2}

1 Leibniz Institut für Präventionsforschung und Epidemiologie – BIPS, Bremen, Deutschland

2 Fachbereich Human- und Gesundheitswissenschaften, Universität Bremen, Deutschland

Einleitung

Zwischen Raum und Gesundheit besteht ein enger Zusammenhang. Jeder Ort bietet unterschiedliche Voraussetzungen und Herausforderungen. So variieren nicht nur sozioökonomische Bedingungen, Umweltexpositionen, Qualität und Charakter der natürlichen und bebauten Umgebung, sondern auch der Zugang zu und die Erreichbarkeit von gesundheitsrelevanten Ressourcen. Gemäß dem *Health in All Policies*-Ansatz sollte die Schaffung gesundheitsförderlicher Lebensverhältnisse vor Ort in allen Politik- und Verwaltungsbereichen berücksichtigt werden. Initiiert wurde dies bereits mit der Ottawa-Charta [1]. Gleichzeitig sollen gesundheitliche Beeinträchtigungen durch gezielte Aktivitäten im Rahmen der Verhältnisprävention verhindert werden. Spezifische Maßnahmen im Sinne des Public Health Action Cycle sind auf datenbasierte Informationen angewiesen. Zwar kann z.B. auch die Gesundheitsberichterstattung (GBE) auf eine Reihe von Datenquellen zurückgreifen, allerdings variieren die Möglichkeiten je nach Raumeinheit und Datenlage [2]. Im Zuge der Digitalisierung sind das Datenangebot, die Zahl der Anbieter und die damit verbundenen Möglichkeiten der Datenverknüpfung (Data Linkage) stark angestiegen. Insbesondere *offene Daten (Open Data, OD)*, die von Behörden, privaten oder staatlichen Unternehmen, Forschungseinrichtungen oder Bürgerinitiativen verfügbar gemacht werden, ermöglichen neue Zugänge und Nutzungsoptionen. So werden aktuell unter dem Stichwort *Open Government Data (OGD)* mehr und mehr Angebotskanäle für Verwaltungsdaten sowie Standards zur Recherche und Bereitstellung erarbeitet und ausgewählte Daten auf hauseigenen Websites und/oder übergeordneten Portalen angeboten. Dennoch liegt Deutschland im internationalen Vergleich bei der Ausschöpfung dieses Potenzials für die Public-Health-Forschung zurück [3]. Eine der Hauptursachen ist der unübersichtliche Angebotsmarkt [4]. So betreiben z.B. nur einige Kommunen und Regionalverbände OD-Portale, die zudem nur teilweise in übergreifende Kataloge eingebunden bzw. untereinander vernetzt sind. Folglich bleibt bereits das Auffinden von Daten, insbesondere für spezielle räumliche Skalierungen und den hier thematisierten Kontext, herausfordernd. Erschwert wird eine Nutzung zusätzlich durch unterschiedliche technische und rechtliche Barrieren. Bezüglich der FAIRifizierung (FAIR steht als englisches Akronym für Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwertbarkeit [5]) von Daten zeigen sich demnach sehr ähnliche Probleme wie bei Forschungsdaten insgesamt.

Zielsetzung und Methodik

Diese Arbeit soll eine Orientierung bezüglich der aktuell für die Betrachtung von Raum und Gesundheit auf verschiedenen administrativen Ebenen zur Verfügung stehenden Datenangebote geben, die barrierefrei, ohne technische und rechtliche Einschränkung genutzt und

weiterverbreitet werden können. Die Auswahl relevanter Determinanten orientiert sich an theoretischen Rahmenmodellen. Da der Gegenstand aktuelle Praxisinitiativen erfassen sollte und für diesen Kontext bislang kaum wissenschaftlicher Output vorliegt, wird die Verfügbarkeit entsprechender Daten anhand einer einfachen Webrecherche (Google) überprüft und durch das Nachverfolgen von Verweisen auf themenrelevanten Websites, Katalogen und Datenportalen ergänzt. Auf der Basis einer aktuellen Bestandsaufnahme werden Datenquellen identifiziert und nach einer definitorischen Einführung zu OD und der Beschreibung des konzeptionellen Rahmens entsprechend theoretischer Dimensionen eingeordnet. Hierfür wurden vorhandene, offen zur Verfügung gestellte Datenübersichten, z.B. in Form von Webvisualisierungen, auf den Themenschwerpunkt angepasst und erweitert. Unter einer offenen Lizenz stehen diese zur Weiternutzung zur Verfügung. Die Informationen können für Politik, Planung und Wissenschaft gleichermaßen von Bedeutung sein, z.B. als Grundlage regionaler präventiver Aktivitäten und epidemiologischer Studien.

Forschungsstand

Der fachübergreifende Diskurs zu OD, insbesondere zur Weiterverwendung, steht in Deutschland noch relativ am Anfang und wird vor allem im Verwaltungskontext der OGD vorangetrieben. Hier liegen wenige, überblicksartige Fachbeiträge [6], [7], [8] oder Web-Visualisierungen [9], [10], [11] vor. Für den Open-Data-Kontext existieren für Deutschland, neben ausschließlich geodatenbasierten Reviews [12], [13], [14] und Übersichtsarbeiten, die sich vorwiegend auf amtliche Datenquellen für die Public-Health-Forschung konzentrieren [2], [15], [16], [17], bislang keine Übersichten für den spezifischen Kontext der raumbezogenen Gesundheitsforschung im Sinne relevanter, offener Datenquellen, deren Zugangswege und Nutzungsmöglichkeiten.

Open Data

Veröffentlichte Daten sind nicht zwangsläufig offen. Beispielsweise dann nicht, wenn für Daten Nutzungsrechte vorbehalten sind, nur fallspezifisch gewährt werden oder wenn die Daten in einem Format vorliegen, das eine Weiternutzung erschwert. Offen per Definition – in technischer und rechtlicher Hinsicht – sind Daten vor allem dann, wenn sie maschinenlesbar und strukturiert sind und unter Verwendung offener Nutzungsrechte von jedermann frei über eine offene Schnittstelle verwendet, nachgenutzt und verbreitet werden können – und zwar zu jeglichem Zweck, ohne Einschränkungen, Diskriminierungen und Gebühren [18]. Eine effiziente Weiternutzung von Daten setzt demnach voraus, dass diese in guter Qualität, als Rohdaten, bestenfalls inklusive Metadaten, vorliegen und klar definierte Nutzungsrechte ausweisen. Nutzungsbedingungen (Lizenzierungen) werden im Rah-

Tabelle 1: Rechtliche und technische Offenheit

| Rechtliche Offenheit: gängigste Lizenzen und Anwendungskontexte | |
|---|--|
| Creative Commons (CC) | Gilt nicht ausschließlich für Daten und erlaubt, Werke unter den angegebenen Bedingungen zu kopieren, zu verbreiten und anderweitig zu nutzen. CC ist weltweit einsetzbar, gilt für die Dauer des Urheberrechtes und wird häufig für Geodaten genutzt. Lizenzgeber können Erlaubnisse hinzufügen, um Bedingungen festzulegen. Demnach unterscheiden sich die Lizenzmodelle in den Versionen 1.0 bis 4.0, wobei Letztere für den internationalen Kontext gilt und z.B. von EU-Datenbereitstellern genutzt wird. https://creativecommons.org/licenses |
| Open Database License (ODC) | Liegt in folgenden Versionen vor: <ul style="list-style-type: none"> • Open Data Commons Public Domain Dedication and License (PDDL): Verzicht auf alle Urheberrechte an Daten • Open Data Commons Attribution License (ODC-BY): Teilen und Bearbeiten erlaubt; Namensnennung erforderlich • Open Data Commons Open Database License (ODC-ODbL): Teilen und Bearbeiten erlaubt; Namensnennung und Weitergabe unter gleichen Bedingungen http://opendatacommons.org/licenses/odbl/ |
| Datenlizenz Deutschland | Hauptsächlich für OGD genutzt, liegt aktuell in folgenden Versionen vor: <ul style="list-style-type: none"> • Datenlizenz Deutschland – Namensnennung – Version 2.0 – verpflichtet zur Nennung des Bereitstellers nach dessen Maßgabe und des Lizenznamens • Datenlizenz Deutschland – Zero – Version 2.0 – ermöglicht eine einschränkunglose Weiterverwendung https://www.govdata.de/web/guest/lizenzen |
| Formate für Texte, Tabellen und Bilder | .txt, .csv, .htm, .xml, .rdf, .odt, .ods, .rss, .pdf, .doc, .docx, .xls, .xlsx, .rtf, .gif, .jpg, .png, .tif, .geotiff |
| Formate für Geodaten | .gml, .gpx, .kml, .dxf, .dwg, .shp, .shx, .dbf, .svg |

kursiv: maschinenlesbar, **fett**: offenes Format, **fett + kursiv**: offen und maschinenlesbar
Eigene Zusammenstellung, technische Offenheit angelehnt an [19]

men der *rechtlichen Offenheit* adressiert. Viele öffentliche Daten können nur unter Beachtung ausgewiesener Bestimmungen, z.B. hinsichtlich jener, die eine kommerzielle Nutzung, Bearbeitung und Weitergabe erlauben, genutzt werden. Für gesetzlich geschützte personenbezogene Daten ist eine Nachnutzung überwiegend nicht gestattet. Eine Orientierung geben die in Tabelle 1 aufgeführten gebräuchlichen Lizenzen, die eine Weiterverarbeitung im OD-Sinn überwiegend uneingeschränkt erlauben. Die *technische Öffnung* betrifft die Datenformate, die eine Weiternutzung sowohl mit proprietärer als auch freier Software gewährleisten. In den letzten Jahren wurden Empfehlungen für die Veröffentlichung von Daten erarbeitet. Die in Tabelle 2 aufgeführten Prinzipien für OGD gelten seither als Standard.

Datenherkunft

Offene Daten, die für die raumbezogene Public-Health-Forschung genutzt werden können, inkludieren neben OGD auch ein breites Spektrum offener Geo-Daten sowie Daten der amtlichen Statistik, die zum Teil bereits mit Geo-Daten verknüpft sind. Auch Unternehmen, Forschungseinrichtungen und die Zivilgesellschaft stellen Daten in offenen Formaten zur Weiternutzung zur Verfügung. Nachfolgend werden die in Abbildung 1 dargestellten Quellen näher beschrieben.

OGD und Fachdaten

Die Bezeichnung ‚OGD‘ verweist auf die Herkunft aus dem öffentlichen Sektor (Behörden, Ämter, staatliche Stellen) [19]. Derartige Daten liegen häufig nur aggregiert vor, sind teils schwer auffindbar oder nur auf Nachfrage (z.B. auf Grundlage des Informationsfreiheitsgesetzes (IFG)) zugänglich. So fertigen zahlreiche Behörden sogenannte Auftragslieferungen an, die zwar meist entgeltfrei sind, demnach definitionsgemäß aber nicht unter die Prämisse von OD fallen. Während OGD bislang überwiegend anbieterbestimmt angeboten wurden, sollen diese künftig vermehrt proaktiv, für jedermann, jegliche Zwecke, ohne Einschränkungen, kostenfrei weiternutzbar und -verbreitbar bereitgestellt werden [20]. *Ob* bzw. *als wie offen* OGD bezeichnet werden können, wird durch die zur Verfügung gestellten Formate und den Grad der Zugänglichkeit determiniert [19]. Zur Quantifizierung dessen und als Qualitätsbeleg nutzen immer mehr Datenbereitsteller das von Berners-Lee vorgeschlagene kaskadierende (den zweiten Stern kann nur bekommen, wer bereits die Voraussetzungen für den ersten Stern erfüllt) 5-Sterne-Modell, dessen Basis eine offene Lizenz voraussetzt und an dessen Spitze im Web veröffentlichte Daten stehen, die in einem strukturierten, nicht proprietären Format mit einer eindeutigen URL zur Datenverlinkung vorliegen [21].

Tabelle 2: Kriterien offener Daten [56]

| Kriterium | Beschreibung der Anforderungen |
|---|---|
| Vollständigkeit | So vollständig wie möglich, soweit dies die Regelungen zum Datenschutz zulassen; möglichst im Rohformat, inklusive beschreibender Metadaten |
| Primärquellen | Aus Ursprungsquellen, mit Informationen zur Erhebung und Aufbereitung |
| Zeitliche Nähe | Veröffentlichung zeitnah zur Entstehung oder Aktualisierung der Daten |
| Leichter Zugang | Vermeidung/Minimierung von Hindernissen, z.B. durch technische Probleme, Dateiformate, fehlende Metadaten, organisatorisch Barrieren (z.B. Pflichtregistrierung), rechtliche Vorgaben |
| Maschinenlesbarkeit | Verwendung einheitlicher, maschinenlesbarer Formate |
| Diskriminierungsfreiheit | Zugriff ohne zeitliche Restriktionen oder notwendige Registrierungen, keine Einschränkung auf ausgewählte Softwareanwendungen |
| Verwendung offener Standards | Verwendung von Formaten, die nicht nur von ausgewählten Programmen gelesen und verarbeitet werden können, keine Notwendigkeit der Abführung von Lizenzkosten an Software-Hersteller |
| Nutzungsbestimmungen/ Lizenzierung | Klare und einheitliche Nutzungsbedingungen, eindeutig sichtbare Lizenz, die Nutzungsvoraussetzungen festlegt |
| Dauerhaftigkeit | Permanente Verfügbarkeit, nachvollziehbare Dokumentation von Änderungen, Aktualisierungen und Löschungen mit Versionskontrolle und Archivierung |
| Nutzungskosten | Keine Bereitstellung gegen Entgelt |

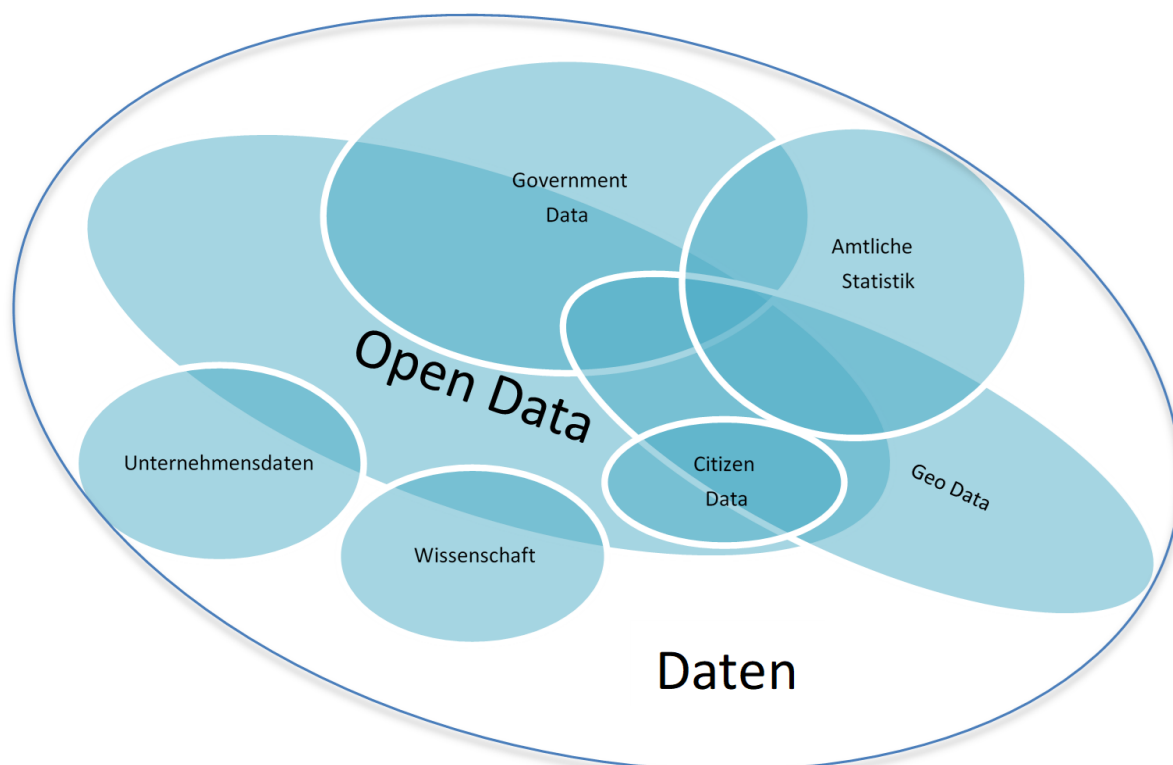


Abbildung 1: Offene Daten und Schnittmengen nach Datenquellen

Nationale Portale für OGD, wie *Open Data Monitor* oder *GDI-DE*, werden seit mehreren Jahren staatlich unterstützt und fassen regionale und kommunale Angebote zusammen. Das nationale Metadatenportal *GovData.de* hält selbst keine Daten vor, verweist aber auf OGD aus allen föderalen Ebenen. Überdies stellen neun Bundesministerien OGD auf hauseigenen Portalen bereit. Das Bundesministerium für Verkehr und digitale Infrastruktur (BMVI) bietet mit dem *mdm-Portal*, *CODE-DE2*, Daten des *Copernicus-Erdbeobachtungsprogramms* sowie der *mCLOUD* (für Mobilitäts-, Geo- und Wetterdaten) mehrere Zugänge. Letztere dient gleichzeitig als Rechercheplattform zu OD privater Anbieter aus dem Mobilitätsbereich. Hinzu kommt der inzwischen entgeltfreie Zugang zu offenen meteorologischen und klimatologischen Fachdaten auf dem *OD-Portal des Deutschen Wetterdienstes* (DWD). Einige Ministerien stellen Daten nicht selbst, aber über Open-Data-Portale bereit.

Offene Geodaten

Offene Geodaten kommen aus Verwaltung, Forschung, freier Wirtschaft, oder von Crowd-Sourced-Initiativen wie *OpenStreetMap* (OSM) und können als thematisch aufbereitete *Geofachdaten* oder liegenschaftsbeschreibende und topographische *Geobasisdaten* verfügbar gemacht werden [22]. Für die Public-Health-Forschung entsteht der Wert dieser Ressource erst durch ein *Data Linkage* mit Sachdaten [23], [24] und damit der Möglichkeit, ökologische Zusammenhänge oder räumliche Disparitäten in Bezug auf gesundheitsrelevante Fragestellungen zu identifizieren [25]. Voraussetzung ist, dass die Daten aktuell, vollständig, qualitätsgeprüft, in einheitlichem Format und interoperabel (z.B. im europaweit harmonisierten Metadatenstandard DCDAT), mit transparenten Nutzungsbedingungen und dokumentierten Metadaten, dauerhaft vorliegen [26]. Die Umsetzung der Infrastructure for Spatial Information in Europe (INSPIRE) hat wesentliche rechtliche Grundlagen hierfür geschaffen. In Deutschland werden diese über eine Geodateninfrastruktur Deutschland (GDI-DE) umgesetzt. Zentrale Recherche-Zugänge für offene Geodaten auf allen föderalen Ebenen bieten *GovData.de* und *Geoportal.de*. Auch das Bundesamt für Kartographie und Geodäsie (BKG) bietet OD an. In Regionaldatenbanken werden kartenfähige Datensätze über interoperable Downloaddienste angeboten. Bei diesen handelt es sich allerdings nicht notwendigerweise immer um OD. Auch private Anbieter wie ESRI und Google (*GoogleMaps*) stellen ausgewählte Geodaten mit freien Lizenzen zur privaten Nutzung bereit, allerdings sind diese teilweise nur mit proprietärer Software nutzbar. Eine weitere, gesundheitsrelevante Raum-Daten-Ressource, bereitgestellt durch den GKV-Spitzenverband und die Deutsche Krankenhausgesellschaft (DKG) befindet sich seit Januar 2020 im Regelbetrieb: Die Webanwendung *krankenhausstandorte.de* ermöglicht es aus der Bereitstellungsperspektive, geocodierte Krankenhaus- und Ambulanzeinrichtungen inklusive Metadaten zu registrieren,

die dann von der Nutzerseite (allerdings erst nach einer Registrierung) abgerufen werden können.

Offene Daten der amtlichen Statistik

Daten der amtlichen Statistiken umfassen sowohl Geos als auch Register-, Sozial-, Umwelt- und Gesundheitsdaten. Auf Bundesebene ist das Statistische Bundesamt der führende Datenbereitsteller. Auch über Regionaldatenbanken, Stadt- oder Kommunalverwaltungen sowie über übergreifende Portale und Datenbanken wie die *INKAR-Datenbank* des BBSR lassen sich amtliche OD beziehen. Auch die Sozialhilfestatistik und die Sozialberichterstattung enthalten relevante OD auf nationaler Ebene, aber zum Teil auch in kleinräumiger Gliederung, auf *Landes- oder Kommunalportalen*. Allerdings werden derartige sogenannte Sozialstrukturatlanten bislang überwiegend im nicht-maschinenlesbaren PDF-Format angeboten.

Offene Unternehmensdaten

Das *Datenportal der Deutschen Bahn AG* ist ein Beispiel eines privaten Anbieters, der auf nationaler Ebene OD in einem eigenen Portal zur Verfügung stellt. Auch die von neun großen deutschen Verkehrsverbänden betriebene Initiative *Open-Data-Plattform ÖPNV* stellt auf ihrem Portal OD zur freien Verfügung. Überdies bieten mehr und mehr regionale und städtische Verkehrsbetriebe offene Daten auf ihren Unternehmenswebseiten (z.B. *NVBW Open Data*) an.

Open Data im Rahmen von Citizen Science

Neben staatlichen Aktivitäten erheben und sammeln Freiwillige (in sogenannten Crowd-Sourced-Initiativen) Daten und stellen diese unter offenen Lizenzen zur freien (Weiter-)Nutzung zur Verfügung. Bekannte Initiativen sind die *Projekte des OK-Labs Stuttgart* mit den Plattformen <https://sensor.community/de/> und *OpenSenseMap.org*, einige Projekte des *Codefor-Labs* (<https://www.codefor.de/projekte>). Das ehemalige, inzwischen stillgelegte Portal *offenedaten.de*, auf dem erstmalig deutschlandweit OD auf Basis von Webcrawler-Treffern interaktiv kartiert wurden, dient bis heute als Prototyp derartiger Anwendungen [13]. Ähnliche Konzepte verfolgen der *OpenDataAtlas* [9] und *Opendata-City-Census*. Besonders deutlich wird das Potenzial dieser Ressourcen bezüglich Aspekten wie Urheberschaft, Nutzungsbedingungen, Kosten und Verfügbarkeit (z.B. von OD, die von offizieller Seite fehlen, nicht einheitlich erfasst bzw. angeboten werden).

Offene Forschungsdaten (Open Science)

Auch offen zur Verfügung gestellte Forschungsdaten können zum Teil weiterverwendet und überdies mit räumlichen Daten verknüpft werden. Förderlich in diesem Kontext ist die zunehmende Selbstverpflichtung zu *FAIR*

Data Principles [5]. Allerdings müssen die Daten, die diese Prinzipien erfüllen, nicht zwangsläufig offen sein. Eine potenzielle Quelle sind auch (Gesundheits-)Surveys. Für Regionalisierungen nationaler Erhebungen müssten die Stichproben jedoch sehr groß werden, weshalb für viele interessierende Aspekte keine brauchbaren flächen-deckenden, kleinräumigen Daten vorliegen. Teilweise beschränken sich Nutzungsmöglichkeiten auf wissenschaftliche Zwecke und werden demnach nur als Scientific Use Files (SUF) angeboten. Ein internationales Register mit detaillierten Informationen zu über 2000 Forschungsdatenspeichern mit offenen Daten zur Weiter-nutzung, unter ausgewiesenen Bedingungen, bietet die Plattform *Registry of Research Data Repositories* (*re3data.org*).

Konzeptioneller Rahmen

Gesunde Lebensbedingungen sind ein mehrdimensionales Konzept mit komplexen Wechselwirkungen. Galt die Aufmerksamkeit in der Vergangenheit vor allem individuellen [27], [28], [29] und umweltbedingten [30], [31], [32] Risikofaktoren, werden aktuell vermehrt protektive Determinanten im Sinne der Salutogenese einbezogen, z.B. soziale Strukturen, Verfügbarkeit von Umweltressourcen (z.B. Grünflächen) sowie potenzielle Faktoren, welche die körperliche Aktivität beeinflussen können [33], [34]. Das ganzheitliche humanökologische Modell [35], das Einflussfaktoren aus sozialen, ökologischen und ökonomischen Systemen aufzeigt, kann zur Strukturierung von Daten herangezogen werden. Die Dimensionen entsprechen den Aspekten, die auch die WHO auf Basis eines Reviews [36] als unerlässlich für das Wohlergehen der Menschen identifiziert hat. Dazu gehören neben individuellen (Lebensstil-)Faktoren auch strukturelle Bedingungen des Lebensumfelds wie (1) soziodemografische und -ökonomische Faktoren, (2) Faktoren des Zugangs und der Erreichbarkeit, (3) die physische Umgebung und (4) Klima- und Umweltfaktoren. Dieser theoretische Rahmen bildet die Basis für die Auswahl und Strukturierung der Daten, wie sie nachfolgend in den Ergebnissen der Daten-Recherche abgebildet sind.

Ergebnisse: Relevante Dimensionen und Datenressourcen

Zwar liegen im Themenfeld *Umfeld und Gesundheit* auf nationaler und kommunaler Ebene teils komplette Datensets, z.B. der Umweltberichterstattung (UBE) sowie den Bund-Länder-GBE-Indikatoren vor, allerdings ohne sich aufeinander zu beziehen [37]. Zudem liegen die Daten überwiegend in nicht-offenen Berichtsformaten vor. Entsprechend der vorab definierten Dimensionen können jedoch die nachfolgend thematisierten und in Anhang 1, Tabelle 3 sowie der zugehörigen Webvisualisierung (<https://sciencemap.github.io/Open-Data-for-Health>)

aufgeführten spezifischen Datenressourcen herangezogen werden.

Indikatoren der sozialen Umwelt: soziale Lage, soziodemografische und sozioökonomische Faktoren

Verschiedene Expositionsniveaus und damit gesundheitliche Effekte werden u.a. durch die sozioökonomische und soziodemografische Komposition der Bevölkerung determiniert [38], [39], [40], wobei die soziale Ungleichheit als Verstärker für Umweltungerechtigkeit fungiert [41]. Für die Epidemiologie werden die Faktoren Bildung, Stellung im Beruf und das Einkommen als wichtigste Merkmale dieser Dimension definiert [17], [42]. Zudem bilden das wirtschaftliche Umfeld, die öffentliche Finanzlage und die Höhe des kollektiven Wohlstands das Sozialkapital und die Leistungsfähigkeit einer Region zur Daseinsvorsorge ab. Eine nützliche Übersicht zu Datenressourcen dieser Dimension, die u.a. die Verwendung von amtlichen Statistiken, der Sozialberichterstattung und GBE ausweist, haben Bardehle et al. [17] erarbeitet. Hinweise auf sozialräumliche Gesundheitsunterschiede können anhand einer Verschneidung mit Geodaten gewonnen werden. Überdies haben sich für Deutschland in der Vergangenheit komprimierte Indikatorensysteme zur Messung sozioökonomischer Deprivation etabliert [43]. Von diesen stehen z.B. die Daten des German Index of Socioeconomic Deprivation (GISD) [44] öffentlich und in maschinenlesbarem Format zur Verfügung, gestatten aber laut Lizenz keine Weiterverarbeitung.

Das Wohlbefinden kann auch vom persönlichen Gefühl der Verwundbarkeit (z.B. durch Kriminalität und Sicherheit im Straßenverkehr) beeinträchtigt werden. Obgleich das Bundeskriminalamt (BKA) seine polizeiliche Kriminalitätsstatistik (PKS) auf Bundesebene seit 2014 nicht mehr nur als PDF-Datei zur Verfügung stellt, fehlen häufig kleinräumige Daten in maschinenlesbarer Form bzw. sie sind nur auf Anfrage verfügbar.

Verfügbarkeit, Zugang und Erreichbarkeit

Der Zugang zu Dienstleistungs- und Versorgungsangeboten betrifft u.a. sowohl die generelle Verfügbarkeit als auch den Zugang im Sinne der Erreichbarkeit. Auch gesellschaftlichen Bedingungen (Partizipation, Teilhabe) und damit vorhandenen sozialen Infrastrukturen kommt aufgrund ihrer salutogenen Schutzfunktion [45] eine hohe Gesundheitsrelevanz zu. Die Lebenserwartung bei Geburt kann als Indikator für die allgemeinen Lebens- und Versorgungsbedingungen vor Ort betrachtet werden. Während Hausärzte eine zentrale Institution in der ambulanten Gesundheitsversorgung darstellen, wurde für die vergangenen Jahre außerdem ein steigender Versorgungsbedarf für bestimmte Fachärzte erkannt [46]. Zudem spielen sowohl die Verfügbarkeit als auch die Erreichbarkeit von Bewegungs- und Sozialisierungsmöglichkeiten

eine zentrale Rolle [47]. Neben OD aus der amtlichen Statistik sind Geodaten eine wertvolle Ressource zur Abbildung von Versorgungsrealitäten auf verschiedenen räumlichen Ebenen.

Physische Umgebung

Die physische Umwelt hat insbesondere auf kleinräumiger Ebene eine hohe Bedeutung für die Gesundheit. Dies betrifft sowohl die mentale als auch die physische und die soziale Komponente [48]. Die Nutzungsmischung, also das Verhältnis von Freiräumen, Verkehrsflächen, Gebäuden und wichtigen Versorgungsstrukturen, ist hierbei eine zentrale Größe. Regionale und örtliche Katasterämter bieten größtenteils Informationen zu öffentlichen Grünflächen als OD an. Auch offen zugängliche Indikatoren wie Bodenbedeckung, Bebauungsdichte und Bodenversiegelung können diese Aspekte abbilden und gleichzeitig Hinweise auf die ökologische Situation und klimarelevante Belastungen geben. So birgt ein hoher Versiegelungsgrad u.U. ein örtliches Überwärmungspotenzial und damit eine Gefahr für Hitzestress. Nicht zuletzt weisen umfangreiche Industrieflächen auf eine gesteigerte Produktion von Abgasen hin [49]. Gebiete mit geringer Bebauungsdichte und Bodenversiegelung weisen hingegen häufig einen hohen Grünanteil auf, der als Schadstofffilter und Sauerstofflieferant dient, die Lufttemperatur senkt und die Luftfeuchtigkeit erhöht. In Bezug auf Frei-, Grün- und Erholungsflächen existieren verschiedene OD-Ressourcen (Anhang 1, Tabelle 3).

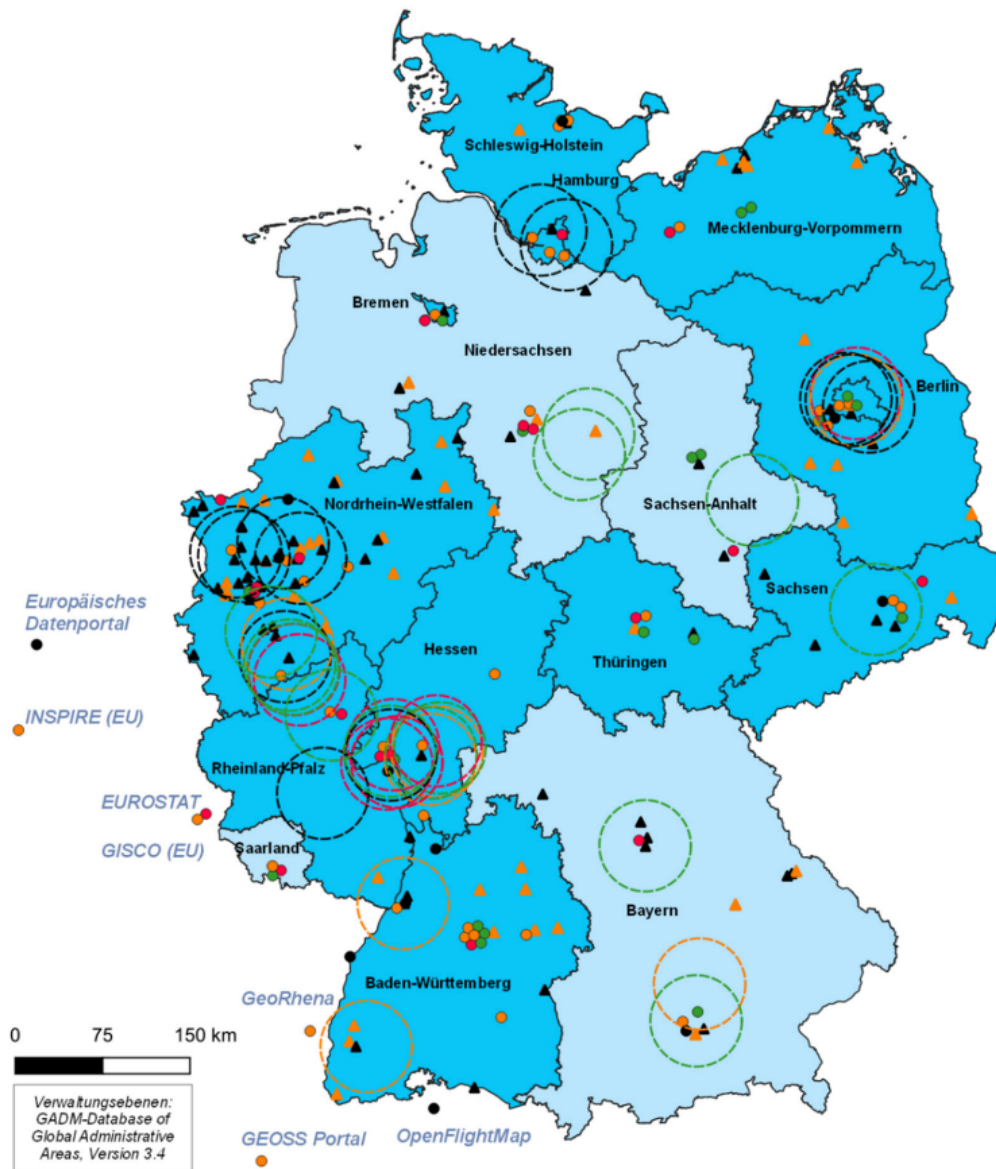
Klima und Umwelt

Umweltrisiken sind physikalische, chemische und biologische Faktoren, die von außen auf Personen einwirken. Von diesen gilt die Luftverschmutzung als Ursache für Krankheiten mit der größten zuschreibbaren Krankheitslast [50]. Zu den anerkannten Indikatoren zur Messung der Belastung gehören die Emission der Stoffe Schwefeldioxid (SO₂), Stickstoffoxide (NO_x), Ammoniak (NH₃) und Feinstaub. Auf Grundlage des Bundesimmissionsschutzgesetzes (BImSchG) werden vor allem in Regionen mit hohem Verkehrsaufkommen (u.a. in Großstädten und Ballungsgebieten) Messstationen zur Ermittlung und Überwachung der Luftexpositionen unterhalten. Informationen zur räumlichen Verteilung derartiger Schadstoffkonzentrationen werden teilweise auf Basis geostatistischer Interpolationen angeboten. Da allerdings die Verkehrsnähe relevant ist, besitzt die jährliche Gesamtemission (in Tonnen) in definierten Verwaltungseinheiten nur eine geringe Aussagekraft. Somit sind Ableitungen für flächendeckende, kleinräumige Expositionen wie z.B. Tageskonzentrationen für Feinstaub nur eingeschränkt möglich. Dies gilt gleichermaßen für andere Daten, die nur an bestimmten (Mess-)Stellen erhoben werden, wie z.B. meteorologische Daten des DWD. Direkte Umwelteinflüsse zeigen sich auch in Form klima- oder umgebungs-lärmbedingter gesundheitlicher Folgen. Lärmkarten zum Straßenverkehr liegen hauptsächlich für Großstädte mit

mehr als 100.000 Einwohnern vor, teilweise sogar als flächendeckende Modellrechnungen und Daten aus Lärmgutachten. Für kleinere Städte und Gemeinden stellt sich die Lage allerdings heterogener dar. Auch klimaökologische Daten (und Klimafunktionskarten) liegen vereinzelt, insbesondere kleinräumig, vor. Daten zur Boden- und Wasserqualität (Faktoren, die indirekt auf die Gesundheit einwirken) sind auf vielen Ebenen überwiegend verfügbar.

Räumliche Ebene und Datenverfügbarkeit

Ogleich es ein politisches und gesellschaftliches Interesse gibt, Gesundheitsförderung vor allem auf kommunaler Ebene zu implementieren, ergeben sich insbesondere für kleinräumige Analysen Probleme. Für kleinste räumliche Einheiten werden aus Datenschutzgründen keine Rohdaten veröffentlicht und stattdessen je nach Verwendungszweck aufbereitet bzw. aggregiert. Zudem erschweren unterschiedliche Gebietsgrenzen und Aggregationen die Vergleichbarkeit und ein Data Linkage mit Sachdaten. Neben dieser eingeschränkten Verfügbarkeit und begrenzten Regionalisierbarkeit wird die Datennutzung vor allem durch föderale Strukturen erschwert. So können zentrale Portale wie *Gov.Data* nur kommunale Datensätze aufnehmen, wenn die zugehörigen Bundesländer (BL) das Portal unterstützen. Allerdings sind bislang nur 12 BL der Vereinbarung beigetreten. Stattdessen haben einige BL eigene Datenportale eingerichtet. Die Folge sind heterogene Datenumfänge, -zugänge und rechtliche Rahmen (teils ohne eigene gesetzliche Regelungen auf Basis des IFG). Auch die kommunale GBE ist länder- oder kommunenspezifisch ausgerichtet und weist keinen gemeinsamen Grundstandard auf. Trotz der langsam steigenden Zahl der Kommunen, die OD veröffentlichen (besonders in Nordrhein-Westfalen, Hamburg und Berlin), ist das Angebot aktuell noch sehr selektiv bzw. in einigen Regionen kaum vorhanden (Abbildung 2). Auf verschiedenen räumlichen Ebenen stellen einige Akteure zum Teil umfangreiche, andere jeweils thematisch nicht vergleichbare oder überhaupt keine (insbesondere in finanzschwachen Kommunen) OD bereit. Von über 11.000 Kommunen in Deutschland veröffentlichten im Jahr 2021 erst rund 146 Kommunen überhaupt Daten, davon mit teils weniger als 10 Datensätzen [51]. Für den hier thematisierten Kontext reduziert sich die Zahl entsprechend weiter. Anhang 1, Tabelle 4 gibt einen Überblick über die Datenressourcen, Datentypen und Lizenzen auf den verschiedenen Verwaltungsebenen, die auf der Grundlage dieser aktuellen Untersuchung ermittelt werden konnten.



National

- ◆ Portal für amtliche Daten
- ◆ Fachdatenportal (z.B. Umwelt, Klima)
- ◆ Geodatenportal
- ◆ Open-Data-Portal oder -Ressource

Kommunal (Städte & Gemeinden)

- ▲ Portal für amtliche Daten
- ▲ Fachdatenportal (z.B. Umwelt, Klima)
- ▲ Geodatenportal
- ▲ Open-Data-Portal oder -Ressource

International & regional (Bundesländer, Metropolregionen)

- Portal für amtliche Daten
- Fachdatenportal (z.B. Umwelt, Klima)
- Geodatenportal
- Open-Data-Portal oder -Ressource
- BL (mit Kreisen & Kommunen) in GOV.data.de enthalten
- BL, die GOV.data.de nicht beigetreten sind

Anmerkung: Abgebildet ist jeweils der geocodierte Standort des Datenbereitstellers. Einige Koordinaten wurden verschoben, um eine Überlagerung in der Visualisierung zu umgehen. Die Symbole für die nationalen Portale (Ringe) bilden nicht die Reichweite ab und wurden ausschließlich aus Darstellungsgründen gewählt. Weitere Informationen inklusive Links zu den genannten Ressourcen sowie eine Übersicht kleinräumiger Portale (Stadtebene) finden sich in der zur Publikation gehörigen Webvisualisierung „Open Data for Health“ <https://sciencemap.github.io/Open-Data-for-Health/>.

Abbildung 2: Verfügbarkeit offener Daten nach räumlicher Ebene und Datenquelle (Stand April 2021)

Diskussion

Für die Entwicklung von Gesundheitsstrategien und ein Monitoring dieser sind die Verfügbarkeit sowie die ressort-übergreifende Nutzung und Verknüpfung von Daten essenziell und entsprechen dem Verständnis von *Health in All Policies*. In diesem Sinne verfolgt die Public-Health-Forschung die Strategie, „Daten für Taten“ sichtbar, zugänglicher und disseminierbar zu machen [52], um u.a. politischen Handlungsbedarf zu identifizieren. Zur Analyse von Zusammenhängen zwischen Raum und Gesundheit haben sich angesichts der voranschreitenden Digitalisierung und damit auch der Quantität, Qualität und Nutzungsmöglichkeiten von Daten, von der lokalen bis zur internationalen Ebene, weitreichende Möglichkeiten eröffnet.

Ein offener Zugang zu Daten bildet hierfür die Voraussetzung. Die aufgezeigten Datenressourcen machen deutlich, dass viele Daten, insbesondere offene Geodaten, prinzipiell zur Verfügung stehen, dies aber das Wissen um diese Zugänge voraussetzt. Zudem bleibt häufig unklar, welche Daten tatsächlich als OD veröffentlicht werden. Überdies ist der jeweils aktuelle Angebotstand von OD nur schwer zu bestimmen, da keine zentrale oder offizielle Sammelstelle für diese Informationen existiert. Die zunehmende Zahl verschiedener Portale bietet aufgrund einiger Restriktionen allenfalls einen Rechercheeinstieg. Erschwert wird dies zusätzlich durch jeweils eigene oder fehlende Informationsfreiheit- und Datenschutzgesetze. Alternative Zugänge über Suchmaschinenabfragen setzen allerdings zunächst eine Indizierung der Daten voraus, z.B. auf Basis technischer Offenheit und ausgestattet mit Metadaten [12]. Die thematisierte technische, rechtliche und inhaltliche Heterogenität der Angebote sowie regulatorische Rahmenbedingungen stehen aktuell einer systematischen Verwendung und damit Ausschöpfung des vollen Potenzials dieser Ressourcen noch deutlich im Wege. So ist nach wie vor eine Vielzahl an Daten grundsätzlich zugänglich, aber nicht in maschinenlesbaren, offenen Formaten abrufbar. Zudem variieren die Menge und die Qualität der verfügbaren Daten. Insbesondere die kleinräumige und flächendeckende Verfügbarkeit ist lückenhaft und stark heterogen. Trotz Initiativen wie INSPIRE sind auch das Auffinden und der Zugang zu Geodaten nach wie vor herausfordernd. Da gesundheitsrelevante räumliche Daten vergleichsweise unterrepräsentiert sind [53], profitiert z.B. auch das Monitoring des Robert-Koch-Instituts bislang kaum von INSPIRE [25].

Die Offenheit von Daten inkludiert auch Aspekte der Partizipation und Weiterverarbeitung. Zwar werden behördliche Datenbereitsteller im Rahmen des E-Government-Gesetzes dazu angehalten, Rohdaten zu veröffentlichen, über die Verwendung entscheiden diese allerdings nach wie vor selbst. Überdies blockiert vielfach auch das Fehlen einer ausdrücklichen Lizenz eine Weiternutzung. Auch das Geodatenzugangsgesetz (GeoZG) fordert eine kostenfreie Bereitstellung, sieht aber eine eingeschränkte, häufig kostenpflichtige Weiterverwendung vor. Somit bleibt auch die Lage bezüglich anfallender Gebühren und

(lizenzrechtlicher) Datenpolitiken für Geodaten heterogen und erschwert eine flächendeckende Zusammenführung von Daten. Open-Citizen-Projekte wie OSM sind in diesem Kontext nicht nur aufgrund des Umfangs an Daten ein Zugewinn. Andererseits weisen diese wiederum Defizite bezüglich der Vollständigkeit, Qualität, Aktualität und Validität auf. Dementgegen sind amtliche Geo-Daten häufig auf administrativen Ebenen verfügbar, die für wissenschaftliche Analysen nicht geeignet sind. So stehen kleinräumige epidemiologische Daten häufig auf Postleitzahlenebene zur Verfügung, die sich mit administrativen Einheiten inkonsistent überlagern. Weitere Probleme administrativer Geodaten, die räumliche Analysen beeinflussen und für alternative Zugänge sprechen, sind Gebietsreformen, Datenaggregationen, Maßstabs- und Zonierungseffekte. Aus Verfügbarkeitsgründen können für flächendeckende oder Ebenen-übergreifende Analysen teilweise sogar Indikatorenwechsel nötig werden.

Ausblick

Ein vollständiger Überblick über OD-Angebote ist derzeit aus verschiedenen Gründen nicht möglich. Da derartige Übersichten allenfalls eine Momentaufnahme in einer sich dynamisch entwickelnden Situation abbilden, ist es sinnvoll, diese kontinuierlich zu erweitern. Erstrebenswert wären Portale/Übersichten, die IT-basiert relevante, bestenfalls standardisierte Datenkataloge aus dezentralen Ressourcen importieren. Metadaten (inklusive technischer und rechtlicher Bedingungen) sind hierfür eine wesentliche Grundlage und setzen den Einsatz von (Meta-)Datenkatalogen bzw. eine eindeutige Benennung von Datensätzen seitens aller Daten-Bereitstellenden voraus. Die Zurverfügungstellung von OD ist ein Gemeinschaftsprozess auf mehreren Mikro- und Makroebenen (Staat, BL, Kreise, Kommunen, Institutionen) und impliziert zudem, dass neben einer dringend notwendigen Harmonisierung rechtlicher Aspekte künftig bereits im (amtlichen) Datenerhebungs- bzw. Forschungskontext generelle Prozesse umgestaltet werden müssen, um initial die Option der Öffnung und maschinellen Lesbarkeit von Daten einzuräumen. Insbesondere die institutionellen bzw. organisationalen Stellen der Qualitätssicherung, des Datenmanagements und des Datenschutzes sollten in derartige Pläne und Prozesse frühzeitig integriert werden. Auch im Hinblick auf die künftige (Weiter-)Nutzung von Forschungsdaten im Zuge der FAIRifizierung müssten die Anstrengungen bezüglich der Vereinbarung von Gemeinschaftsstandards, z.B. Ontologien und Fachvokabular, fachspezifisch vorangetrieben werden, um die Auffindbarkeit und Indizierung von Daten- und Metadatenätzen in Fachdatenbanken, aber auch durch übergreifende Suchmaschinen, deutlich zu optimieren. Um die Leistung Letzterer zu verbessern, entwickelten Google, Yahoo, Yandex und Bing bereits den Metadatenstandard *schema.org*, der Datenproduzierende und -Bereitstellende dabei unterstützt, Daten entsprechend eines einheitlichen Vokabulars zu strukturieren. Auch die vermehrte Integration von Linked-

Data-Standards im Bereitstellungsprozess kann künftig die Auffindbarkeit, den Zugang und die Vergleichbarkeit von Daten erheblich verbessern. Für eine potente Weiterentwicklung von Daten sind (harmonisierte) Gesetzesgrundlagen auf Ebene der BL und Kommunen nötig. Eine aktuelle Orientierung für OGD im kommunalen Bereich bietet der Musterdatenkatalog, der Art und Quelle der OGD ausweist [6].

Im Rahmen der OGD hätte der Gesetzentwurf vom Februar 2021 zum *Zweiten Open Data-Gesetz und Datennutzungsgesetz* (§ 12a EGovG) [54] ein wichtiger Schritt sein können. Obgleich dessen Verabschiedung (vom 24. Juni 2021) Behörden der mittelbaren Bundesverwaltung nunmehr dazu verpflichtet, unbearbeitete, maschinenlesbare Daten zum Abruf bereitzustellen, wird der Anspruch auf eben jene Bereitstellung (auch für nicht geschützte, personenbezogene Daten) nach wie vor eingeschränkt. Auch die neu im Gesetz berücksichtigten Daten von Privatunternehmen (u.a. vom ÖPNV) sowie öffentlich finanzierte Forschungsdaten betreffen teils weitreichende, nicht OD-konforme Einschränkungen. So können die anbietenden Stellen beispielweise Gebühren für ihre Daten verlangen und über die Lizenzierung (und damit über die Weiterentwicklung) entscheiden. Eine übergreifende Aufgabe muss es demnach bleiben, generelle Anreize für OD zu schaffen – nicht zuletzt hinsichtlich finanzieller und personeller Ressourcen. In der Public-Health-Forschungslandschaft sollten die wachsenden Möglichkeiten der gemeinsamen Datennutzung, trotz der (Anfangs-)Kosten der FAIRifizierung und der Öffnung von Daten im Sinne der OD-Prinzipien, wahrgenommen, aktiv nachgefragt und genutzt werden.

Abkürzungen

- BBSR: Bundesinstitut für Bau-, Stadt- und Raumforschung
- BImSchG: Bundesimmissionsschutzgesetz
- BKA: Bundeskriminalamt
- BKG: Bundesamt für Kartographie und Geodäsie
- BL: Bundesländer
- BMVI: Bundesministerium für Verkehr und digitale Infrastruktur
- CC: Creative Commons
- DKG: Deutsche Krankenhausgesellschaft
- DWD: Deutscher Wetterdienst
- GBE: Gesundheitsberichterstattung
- GDI-DE: Geodateninfrastruktur Deutschland
- GeoZG: Geodatenzugangsgesetz
- GISD: German Index of Socioeconomic Deprivation
- IFG: Informationsfreiheitsgesetz
- INKAR: Indikatoren und Karten zur Raum- und Stadtentwicklung
- INSPIRE: Infrastructure for Spatial Information in Europe
- OD: Open Data
- ODbL: Open Database License
- ODC: Open Data Commons
- ÖPNV: Öffentlicher Personennahverkehr

- OGD: Open Government Data (offene Verwaltungsdaten)
- OSM: OpenStreetMaps
- PDDL: Public Domain Dedication and License
- PKS: Polizeiliche Kriminalitätsstatistik
- SUF: Scientific Use Files
- UBE: Umweltberichterstattung
- WHO: World Health Organization

Verweise zu den im Text aufgeführten Webressourcen

- CODE-DE2: <https://code-de.org/de>
- Datenportal der Deutsche Bahn AG: <https://data.deutschebahn.com/>
- ESRI Open Data: <https://opendata-esri-de.opendata.arcgis.com>
- Google Maps: <https://www.google.com/intl/de/permissions/geoguidelines/>
- INKAR: <https://www.inkar.de/>
- Krankenhausstandorte: <https://krankenhausstandorte.de>
- mCLOUD: <https://www.mcloud.de/>
- mdm-Portal: <https://www.mdm-portal.de/>
- NVBW (Nahverkehrsgesellschaft Baden-Württemberg mbH) Open Data: <https://www.nvbw.de/open-data>
- OD-Portal des Deutschen Wetterdienstes: <https://www.dwd.de/DE/leistungen/opendata/opendata.html>
- OpenDataAtlas: <http://opendata.tursics.de/>
- Opendata-City-Census: <http://de-city.census.okfn.org>
- Open Data for Health Map: <https://sciencemap.github.io/Open-Data-for-Health/>
- OpenData-Plattform ÖPNV: <https://www.opendata-oepnv.de>
- OpenSenseMap: <https://OpenSenseMap.org>
- OpenStreetMaps: <https://www.openstreetmap.de/>
- Regionale Sozialhilfestatistik (Beispiele): <https://statistikportal.thueringen.de/thonsa/SSDstart.php> und <http://www.gsi-berlin.info/>
- Registry of Research Data Repositories: <http://re3data.org>
- Sensor Community: <https://sensor.community/de/>
- Sozialhilfestatistik national: <https://www.statistikportal.de/de/sbe>

Daten

Daten für diesen Artikel sind im Dryad-Repositorium verfügbar unter: <https://doi.org/10.5061/dryad.djh9w0w1f> [55].

Anmerkungen

Danksagung

Die Autoren möchten sich bei NFDI4Health für die Inspiration und den Wissensaustausch, der für den gesamten Arbeitsprozess sehr wertvoll war, bedanken.

Interessenkonflikt

Die Autoren geben an, dass kein Interessenkonflikt besteht.

Anhänge

Verfügbar unter <https://doi.org/10.3205/000303>

1. Anhang1_000303.pdf (157 KB)
Verfügbarkeit von Daten im Kontext von Raum und Gesundheit, Übersicht Datenzugänge

Literatur

1. Eriksson M, Lindström B. A salutogenic interpretation of the Ottawa Charter. *Health Promot Int.* 2008 Jun;23(2):190-9. DOI: 10.1093/heapro/dan014
2. Trojan A. Integrierte Gesundheitsberichterstattung auf Landes- und kommunaler Ebene: Initiativen und Ansätze der letzten 20 Jahre [Integrated health reporting at the communal and federal state level-policy initiatives and approaches of the last 20 years]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2020 Sep;63(9):1084-93. DOI: 10.1007/s00103-020-03196-7
3. Burgun A, Bernal-Delgado E, Kuchinke W, van Staa T, Cunningham J, Lettieri E, Mazzali C, Oksen D, Estupiñan F, Barone A, Chène G. Health Data for Public Health: Towards New Ways of Combining Data Sources to Support Research Efforts in Europe. *Yearb Med Inform.* 2017 Aug;26(1):235-40. DOI: 10.15265/IY-2017-034
4. Kubicek H, Lippa B. Open Data: Appelle reichen nicht. *Kommune21.* 2015 Mar;15(5):12-3. Available from: https://www.kommune21.de/meldung_21279.html
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJ, Groth P, Goble C, Grethe JS, Heringa J, 't Hoen PA, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016 Mar;3:160018. DOI: 10.1038/sdata.2016.18
6. Krabina B, Wiedemann M. Open Data in Kommunen. Welche Daten stellen Kommunen als Open Data zur Verfügung? Ein Musterdatenkatalog für Nordrhein-Westfalen. Gütersloh: Bertelsmann Stiftung; 2019. (Analysen und Konzepte – LebensWerte Kommune; 2/2019). Available from: https://www.bertelsmann-stiftung.de/fileadmin/files/Projekte/Smart_Country/AK_2_2019_05_08_ONLINE.pdf
7. Begleitforschung Smart Service Welt II, editor. Open Public Data in Deutschland. Rahmenbedingungen und Potenziale der Bereitstellung und Nutzung von Daten des öffentlichen Sektors. Berlin: Institut für Innovation und Technik; 2020. Available from: https://www.iit-berlin.de/iit-docs/e76a033d9c274cf1a65efab271552894_SSW_Open_Public_Data_in_Deutschland2.pdf
8. Knetsch G, Fock J. Open Government Data – Zugang zu Umwelt- und Gesundheitsdaten. UMID. 2013;(3):20-4. Available from: https://www.umweltbundesamt.de/sites/default/files/medien/360/publikationen/open_government_data_s_20_24n.pdf
9. Tursics T. Open Data Atlas. [last accessed 2021 Mar 29]. Available from: <https://opendata.tursics.de/>
10. OpenGeoEdu. OpenDataPortal. [last accessed 2021 Mar 29]. Available from: <https://portal.opengeoedu.de/>
11. Open Knowledge Foundation Deutschland e.V. Offene Daten. [last accessed 2021 Mar 29]. Available from: <https://okfn.de/en/projekte/offenedaten/>
12. Paderta D. Open Data – Raumbezogene Daten. Köln: Gesis; 2012. Available from: https://www.ssoar.info/ssoar/bitstream/handle/document/36474/ssoar-2012-paderta-Open_Data_-_Raumbezogene_Daten.pdf
13. Hinz M, Bill R. Ein zentraler Einstiegspunkt für die Suche nach offenen Geodaten im deutschsprachigen Raum. *AGIT J Angew Geoinf.* 2018;(4):298-307. DOI: 10.14627/537647038
14. Kalasek R, Weninger K. Open GeoData. *Oeff Sekt.* 2015 Apr;41(1):17-28. DOI: 10.34749/oes.2015.1274
15. Fehr A, Lange C, Fuchs J, Neuhauser H, Schmitz R. Gesundheitsmonitoring und Gesundheitsindikatoren in Europa. *J Health Monit.* 2017 Mar;2(1):3-23. DOI: 10.17886/RKI-GBE-2017-004.2
16. Bardehle D, Anuss R, Hermann S, Ziese T, Böhm K. Der neue Länderindikatorenset für die Gesundheitsberichterstattung [The new indicator set for health reporting activities in the German States]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2004 Aug;47(8):762-70. DOI: 10.1007/s00103-004-0872-x
17. Bardehle D, Blettner M, Laaser U. Gesundheits- und soziodemographische (sozialepidemiologische) Indikatoren in der Gesundheits- und Sozialberichterstattung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2001 Apr;44(4):382-93. DOI: 10.1007/s001030050456
18. Open Knowledge Foundation Deutschland. Das ABC der Offenheit. Berlin: OKFN/Wikimedia; 2019. Available from: https://upload.wikimedia.org/wikipedia/commons/a/a9/ABC_der_Offenheit_-_Broschüre_%282019%29.pdf
19. von Lucke J, Geiger C. Open Government Data – Frei verfügbare Daten des öffentlichen Sektors. In: von Lucke J, editor. Entdeckung, Erkundung und Entwicklung 2.0: Open Government, Open Government Data und Open Budget 2.0. Friedrichshafen: Band 1 der Schriftenreihe des Deutsche Telekom Institute for Connected Cities, TICC der Zeppelin Universität; 2012. p. 105-56.
20. Internet und Gesellschaft Collaboratory. Offene Staatskunst – Bessere Politik durch „Open Government“? 1. Auflage; 2010. Available from: <http://d-nb.info/1120152208/34>
21. Janowicz K, Hitzler P, Adams B, Kolas, D, Vardemann C. Five stars of linked data vocabulary use. *Semantic Web.* 2014;5(3):1-3.
22. Meinel G. Grundlagen der Geo-Informationssysteme. *Raumforsch Raumordn.* 2016 Dec;74(6):593-4. DOI: 10.1007/s13147-016-0418-3

23. Fletcher-Lartey SM, Caprarelli G. Application of GIS technology in public health: successes and challenges. *Parasitology*. 2016 Apr;143(4):401-15. DOI: 10.1017/S0031182015001869
24. Nykiforuk CI, Flaman LM. Geographic information systems (GIS) for Health Promotion and Public Health: a review. *Health Promot Pract*. 2011 Jan;12(1):63-73. DOI: 10.1177/1524839909334624
25. Thißen M, Niemann H, Varnaccia G, Rommel A, Teti A, Butschalowsky H, Manz K, Finger JD, Kroll LE, Ziese T. Welches Potenzial haben Geoinformationssysteme für das bevölkerungsweite Gesundheitsmonitoring in Deutschland? Perspektiven und Herausforderungen für das Gesundheitsmonitoring am Robert Koch-Institut [What potential do geographic information systems have for population-wide health monitoring in Germany? Perspectives and challenges for the health monitoring of the Robert Koch Institute]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2017 Dec;60(12):1440-52. DOI: 10.1007/s00103-017-2652-4
26. Kany C, Klein U, Osterhold M, Riecken J, Sandmann S, Schaffert M, Schön B, Seuß R. Wert von Geoinformation. *Z Geodäsie Geoinf Landmanag*. 2018;143(6):390-7. DOI: 10.12902/zfv-0238-2018
27. Kaplan MS, Newsom JT, McFarland BH, Lu L. Demographic and psychosocial correlates of physical activity in late life. *Am J Prev Med*. 2001 Nov;21(4):306-12. DOI: 10.1016/s0749-3797(01)00364-6
28. Lim K, Taylor L. Factors associated with physical activity among older people – a population-based study. *Prev Med*. 2005 Jan;40(1):33-40. DOI: 10.1016/j.ypmed.2004.04.046
29. Sun F, Norman IJ, While AE. Physical activity in older people: a systematic review. *BMC Public Health*. 2013 May;13:449. DOI: 10.1186/1471-2458-13-449
30. Stokols D, Grzywacz JG, McMahan S, Phillips K. Increasing the health promotive capacity of human environments. *Am J Health Promot*. 2003 Sep-Oct;18(1):4-13. DOI: 10.4278/0890-1171-18.1.4
31. Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. *Annu Rev Public Health*. 2006 Apr;27:297-322. DOI: 10.1146/annurev.publhealth.27.021405.102100
32. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. *Am Psychol*. 1992 Jan;47(1):6-22. DOI: 10.1037//0003-066x.47.1.6
33. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens B, Bachmann W. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *J Am Plann Assoc*. 2006;72(1):75-87. DOI: 10.1080/01944360608976725
34. Giles-Corti B, Vernez-Moudon A, Reis R, Turrell G, Dannenberg AL, Badland H, Foster S, Lowe M, Sallis JF, Stevenson M, Owen N. City planning and population health: a global challenge. *Lancet*. 2016 Dec;388(10062):2912-24. DOI: 10.1016/S0140-6736(16)30066-6
35. Barton H, Grant M. A health map for the local human habitat. *J R Soc Promot Health*. 2006 Nov;126(6):252-3. DOI: 10.1177/1466424006070466
36. Neira M, Prüss-Ustün A. Preventing disease through healthy environments: A global assessment of the environmental burden of disease. *Toxicol Lett*. 2016 Oct;259:S1. DOI: 10.1016/j.toxlet.2016.07.028
37. Tobollik M, Kabel C, Mekel O, Hornberg C, Plaß D. Übersicht zu Indikatoren im Kontext Umwelt und Gesundheit [Overview of indicators in the context of environment and health]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2018 Jun;61(6):710-9. DOI: 10.1007/s00103-018-2743-x
38. Hofmeister C, Maier W, Mielck A, Stahl L, Breckenkamp J, Razum O. Regionale Deprivation in Deutschland: Bundesweite Analyse des Zusammenhangs mit Mortalität unter Verwendung des ‚German Index of Multiple Deprivation (GIMD)‘ [Regional Deprivation in Germany: Nation-wide Analysis of its Association with Mortality Using the German Index of Multiple Deprivation (GIMD)]. *Gesundheitswesen*. 2016 Jan;78(1):42-8. DOI: 10.1055/s-0034-1390421
39. Flacke J, Schüle SA, Köckler H, Bolte G. Mapping Environmental Inequalities Relevant for Health for Informing Urban Planning Interventions – A Case Study in the City of Dortmund, Germany. *Int J Environ Res Public Health*. 2016 Jul;13(7):711. DOI: 10.3390/ijerph13070711
40. Briggs D. Environmental pollution and the global burden of disease. *Br Med Bull*. 2003 Dec;68(1):1-24. DOI: 10.1093/bmb/ldg019
41. Bolte G, Bunge C, Hornberg C, Köckler H. Umweltgerechtigkeit als Ansatz zur Verringerung sozialer Ungleichheiten bei Umwelt und Gesundheit [Environmental justice as an approach to tackle environmental health inequalities]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2018 Jun;61(6):674-83. DOI: 10.1007/s00103-018-2739-6
42. Ahrens W, Bellach BM, Jöckle KH. Messung soziodemographischer Merkmale in der Epidemiologie. München: MMV Medizin Verlag; 1998. (RKI-Schriften; 1/98).
43. Maier W. Indizes Multipler Deprivation zur Analyse regionaler Gesundheitsunterschiede in Deutschland: Erfahrungen aus Epidemiologie und Versorgungsforschung [Indices of Multiple Deprivation for the analysis of regional health disparities in Germany: Experiences from epidemiology and healthcare research]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2017 Dec;60(12):1403-12. DOI: 10.1007/s00103-017-2646-2
44. Kroll LE, Schumann M, Hoebel J, Lampert T. Regionale Unterschiede in der Gesundheit – Entwicklung eines sozioökonomischen Deprivationsindex für Deutschland. *Journal of Health Monitoring*. 2017;2(2):103-20. DOI: 10.17886/RKI-GBE-2017-035
45. Gruenewald TL, Karlamangla AS, Greendale GA, Singer BH, Seeman TE. Feelings of Usefulness to Others, Disability, and Mortality in Older Adults: the MacArthur Study of Successful Aging. *J Gerontol B Psychol Sci Soc Sci*. 2007 Jan;62(1):P28-37. DOI: 10.1093/geronb/62.1.p28
46. Walter U, Schneider N, Bisson S. Krankheitslast und Gesundheit im Alter. Herausforderungen für die Prävention und gesundheitliche Versorgung [Morbidity and health in old age. A challenge for prevention and health care]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*. 2006 Jun;49(6):537-46. DOI: 10.1007/s00103-006-1267-y
47. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW; Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not? *Lancet*. 2012 Jul;380(9838):258-71. DOI: 10.1016/S0140-6736(12)60735-1
48. Claßen T. Urbane Grün und Freiräume-Ressourcen einer gesundheitsförderlichen Stadtentwicklung. In: Baumgart S, Köckler H, Ritzinger A, Rüdiger A, editors. *Planung für gesundheitsfördernde Städte*. Hannover: Verlag der ARL; 2018. p. 297-313.
49. Hartig T, Evans GW, Jamner LD, Davis DS, Gärling T. Tracking restoration in natural and urban field settings. *J Environ Psychol*. 2003 Jun;23(2):109-23. DOI: 10.1016/S0272-4944(02)00109-3

50. United States Environmental Protection Agency. Integrated Science Assessment for Oxides for Nitrogen – Health Criteria. Durham, NC: EPA; 2016. Available from: http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=526855
51. Bertelsmann Stiftung. Musterdatenkatalog schafft erstmals Übersicht über offene Daten in Deutschlands Kommunen. 2021 Mar 3. Available from: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Presse/210303_Pressemitteilung_Musterdatenkatalog_schafft_erstmals_Uebersicht_ueber_offene_Daten_in_Deutschlands_Kommunen.pdf
52. Zens M, Shajanian Zarneh Y, Dolle J, De Bock F. Digital Public Health – Hebel für Capacity Building in der kommunalen Gesundheitsförderung: Ausgangslage, Entwicklungsfragen, TEAviisari als modellhafte Implementierung [Digital public health-leverage for community capacity building in health promotion: Current situation, developmental issues and TEAviisari as a model implementation]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz. 2020 Jun;63(6):729-40. DOI: 10.1007/s00103-020-03148-1
53. Seiler M. INSPIRE-Daten im Gesundheitswesen – Status und Potenzial. In: Strobl J, Blaschke T, Griesebner G, Zagel B, editors. Angewandte Geoinformatik. Berlin/Offenbach: Herbert Wichmann Verlag; 2014. p. 303-8.
54. Bundesministerium der Justiz. Gesetz zur Förderung der elektronischen Verwaltung (E-Government-Gesetz - EGovG) – § 12a Offene Daten des Bundes, Verordnungsermächtigung. [last accessed 2022 Mar 1]. Available from: https://www.gesetze-im-internet.de/egovg/_12a.html
55. Peters M, Zeeb H. Data from: Availability of open data for spatial public health research. Dryad Digital Repository. 2022. DOI: 10.5061/dryad.djh9w0w1f
56. Sunlight Foundation. Ten Principles for Opening Up Government Information. 2010 Aug 11. Available from: <https://sunlightfoundation.com/wp-content/uploads/sites/2/2016/11/Ten-Principles-for-Opening-Up-Government-Data.pdf>

Korrespondenzadresse:

Manuela Peters
Leibniz Institut für Präventionsforschung und Epidemiologie – BIPS, Achterstraße 30, 28215 Bremen, Deutschland, Tel.: +49 218-56924
mpeters@leibniz-bips.de

Bitte zitieren als

Peters M, Zeeb H. Availability of open data for spatial public health research. *GMS Ger Med Sci.* 2022;20:Doc01.

DOI: 10.3205/000303, URN: urn:nbn:de:0183-0003030

Artikel online frei zugänglich unter

<https://doi.org/10.3205/000303>

Eingereicht: 19.04.2021

Überarbeitet: 11.11.2021

Veröffentlicht: 04.03.2022

Copyright

©2022 Peters et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.