

An analysis of open data as a location factor

Sebastian Block^{1,*}, and Neele Hiemesch-Hartmann¹

¹ Berlin School of Economics and Law, Department of Business and Economics, Badensche Straße 52, 10825 Berlin, Germany

Abstract

Research background: The publication of data in Germany introduces a paradigm shift in public administrations. The data silos of the administrations are thus being opened up piece by piece. There are many possible applications for these data. Optimized location determination for trade, production, and services is one of these many possible applications. It has been shown that the provision of open data can achieve economic effects and set value chains in motion. The value creation effects can arise directly, indirectly, or downstream. Public data can thus become a relevant location factor for regions and cities.

Purpose of the article: The various implications of opening up data in context with location determination will be examined in this article. This will be done from an economic perspective, highlighting the positive externalities of open data.

Methods: The published data of administrations will be analyzed based on a German data bank of 12,020 datasets concerning their relevance for determining location. For this purpose, hard and soft location criteria will be defined in advance.

Findings & Value added: The result of the study will be a newly developed publication strategy for German public administrations on location criteria. With the help of this publication strategy, open data should function as a location advantage and initiate value creation effects.

Keywords: *open data; open data portals; location factors; digitalization*

JEL Classification: *R32; R50; R58; C82; L17*

* Corresponding author: sebastian.block@hwr-berlin.de

1 Introduction

In recent decades, Germany as a business location has had to contend with production facilities abroad. Lower wage levels abroad made moving labor-intensive production facilities outside Germany attractive for many companies. This development led to a loss of jobs. This shift became evident, especially in regions where a single company is responsible for many jobs (Etezadzadeh, 2021 & Conway, 2011). Stopping or counteracting this development is a challenge for Germany as a business location. The Berlin-Brandenburg region has succeeded in meeting this challenge with Tesla's "Gigafactory" construction project in Grünheide. In the first expansion phase, around 12,000 new jobs are created in the region. These new job offers will result in the need for additional living space in the region. In addition, the region will be better integrated into the transportation infrastructure in Berlin-Brandenburg. The implications of the "Gigafactory" construction project for the Berlin-Brandenburg region in general and the municipality of Grünheide, in particular, are manifold. In addition to expected increased tax revenues, the region's purchasing power will be strengthened, and trade and service offerings will increase. The district administrator of Grünheide describes the "Gigafactory" construction project as a wealth generator for the region (König, 2021).

It is thus clear that a settling economy stimulates a region in many ways. It also becomes clear that regions in Germany compete with each other for the establishment of industrial production facilities. Investors decide for or against a region on the basis of critical figures. Individual hard location factors can only be changed to a limited extent for regions in a competitive situation. Soft location factors, on the other hand, can. These location factors increase the attractiveness of a region. Open Data bundles many relevant location factors and makes rural regions visible to investors. Therefore, the following study examines the Open Data content based on a model region in North Rhine-Westphalia.

2 Theoretical framework

Open Data

First, the term Open Data must be explained. Lucke (2012) defines Open Data as "(...) *all data sets that are made freely available for free use, for further dissemination and free further use in the interest of the general public of society without any restriction*" (Lucke, 2012). Open Data is also seen as part of the Fourth Industrial Revolution in some countries (Khurshid et al., 2019). Open data thus refers to data that can be used without restriction and where public agencies on national, regional and local levels (Begany and Gil-Garcia, 2021) share data for innovation and transparency (Runeson et al., 2021). Further, the unrestricted use of Open Data does not exclude anyone. Public institutions, private companies, and private individuals can use it equally (European Commission, 2020). Example data can be geodata (Liao et al., 2021), machine-readable administratively collected data sets (Mergel et al., 2018), data on metropolitan areas and rural regions, social data, traffic statistics, legal data, administrative data, statistics, research data, scientific publications, or cultural data (Preishe, 2014). Open Data can generate added value in the social, economic, and ecological sense for a region. These added values generate when value creation is in motion. It has been estimated that total productivity gains of effective government data utilization in the economy could be as high as 5%–10% (Leviakangas and Molarius, 2020) but the actual open data use has not yet reached the anticipated levels (Lassinantti et al., 2019). To this end, the administration, here the provider of Open Data, and the user perform specific value creation steps (Nikiforova and McBride, 2021). On the part of the providing administration, these steps range from collecting raw data to the selection, filtering, and cleansing of data sets to the provision and regular updating of the data in an Open Data portal. Here, the user-side steps

of the value chain include data processing and data exploration, transformation, and modeling to synthesize and extract useful and hidden information with high monetary or societal value. After being processed by user groups, the curated datasets are ideally made available. The value creation potential of Open Data is given in different studies with wildly divergent figures. For example, values between EUR 12.1 billion and EUR 131 billion are calculated for value creation potential in the next ten years in Europe (Klenk et al., 2020). The added value that arises in the value chain is then direct, indirect, or subsequent. Directly added value shows a correlation between published data and plausible impact. Thus, data and impact are directly interrelated. For example, labor market data can show that jobs have been created through new databased business models. Indirect added value, on the other hand, can only be imputed indirectly. A correlation between data and impact is argumentative. A direct attribution is only partially evident and can only be completed by assumptions and with deviations. For example, better job placement can be assumed based on detailed and sound labor market data. The subsequently added value does not show any correlation between data and effects. This added value is where welfare effects kick in. This effect means that only under optimistic assumptions, argumentatively, a correlation emerges. For example, the assumption can be made that optimized job placement leads to higher job satisfaction (Dapp et al., 2016).

Location factors

In Germany, 3,288,306 companies are generating a total turnover of EUR 6,770.83 billion in 2019 (Statista, 2021a). In 2020 alone, 542,200 companies were newly founded. These companies need lucrative locations so that they can run their business. In contrast, 10,796 municipalities are vying for companies willing to invest. In total, the municipalities collected EUR 114.9 billion in taxes in 2019, of which EUR 42.63 billion came from trade tax. This amount corresponds to a share of approximately 37 % and shows the relevance of how essential companies are for municipalities (Statista, 2021b). In business practice, a distinction is made between soft and hard location factors when selecting a site. Hard location factors are described as quantifiable characteristics and are directly relevant for the company and more or less directly cost-effective. A distinction is made between the following seven hard factors: The first factor mentioned in the supply of space. Here, the size and location of the land and how they are tailored are critical. In particular, large industrial companies need much space to produce their goods and sufficient storage capacity. Furthermore, the price of the land plays an essential role. This price is an essential factor, especially for companies that are not particularly well endowed financially. The second factor to be mentioned is the transport connection. Here, a sufficiently developed road and rail network and the connection to a waterway and air transport network play an important role. The third factor relates to the geographic location and proximity to suppliers and sales markets. On the one hand, companies have to procure raw materials from supply markets, and on the other hand, they have to sell processed and produced goods and merchandise. The fourth factor involves the labor market. Regions are well advised to offer a balance of skilled labor. This includes, above all, the qualifications as well as the wage level in the respective region. The fifth factor to consider is energy and environmental costs. Here, companies need to check whether a municipality has already made investments in energy and the environment so that the company's energy and environmental costs are tolerable. Here, investments in solar, wind, and water energy can be listed as examples. The penultimate factor to be considered is local taxes and duties. Municipalities are always in competition with other municipalities to attract companies. These potential companies that want to locate in a municipality will carefully consider whether they will be taxed with the 200% minimum or the 400% maximum statutory rate for business tax. The last and seventh factor concerns the subsidy offers, including development loans, investment grants, and subsidies offered by the municipality to

companies to attract them to the region (Meyer-Stamer, 2013). In addition to the classic hard factors already mentioned, there are also soft location factors. These can be divided into company-related and person-related factors. Although the latter is only indirect for the company, they should not be underestimated by the local authorities. Companies attach great importance to providing their employees with a certain quality regarding their living environment. In addition to the environmental quality of the location, this also includes leisure and training facilities, schools and childcare, and an excellent residential environment. In terms of business-related soft factors, the municipality can boast that it has a high standard, especially in the administrative and advertising areas. Before opening a location in a new municipality, companies ask themselves the following: How good is the quality of the labor administration and the image and appearance of the local commercial and industrial areas? What networks, such as associations or support services, does the municipality offer us as a company? The cultural image, its promotion, and the image on the part of the municipality are also very important for the image of some companies (Godau, 2011).

Location marketing

Location marketing promotes a region. The target groups are potential new citizens on the one hand and potential new investors on the other. Location marketing is based on a stringent overall strategy with a clear concept for the aim, target groups, and positioning. Different target groups are addressed uniformly, and reliable facts (for example from Open Data sources) serve as the basis for location marketing. In addition, location marketing is creative, contemporary, and unique in its implementation and is realistic in terms of budget and other resources. However, location marketing in Germany suffers from insufficient budgets that are not used efficiently enough due to a lack of conceptual strength. In Germany, municipalities and regions are already competing for citizens and investments within the European framework (due to the long external borders). On the other hand, urban areas compete with nearby rural regions. Federal states also compete with each other. The location marketing of a region tries to positively influence this competitive situation (Lennardt and Stakemeier, 2017).

3 Methods

For the first time, the sample data catalog offers the opportunity to analyze Open Data structures of different cities, municipalities, and counties in Germany (cf. Wang and Shepherd, 2020). Previously, an analysis problem existed because data sets are named very differently from administration to administration, making a global data analysis almost impossible. The sample data catalog, on the other hand, uses a consistent form of topic and designation. This consistent form makes a systematic analysis possible for the first time. The sample data catalog data set comprises 12,020 data records with the cut-off date being August 31st, 2021 (Bertelsmann Stiftung, 2021). The data set is examined based on the hard and soft company-related and the soft person-related location factors defined at the beginning of this article. For this purpose, the data set of the sample data catalog is supplemented by these three location factor categories and evaluated. In addition, the ten most extensive data focal points are examined according to the responsible administrations. The investigation of the published Open Data will answer the following research question: Which Open Data (location factors) are published on the part of the municipalities and which Open Data (location factors) have not been published so far in order to increase the own location attractiveness?

4 Results

General description of the published data

The 12,020 data sets are from the administrations of 176 cities, local municipalities, and municipal associations. The largest share of the data sets is accounted for by the administrations of the cities of Hamburg (18.29 %; n = 2,198), Berlin (9.97 %; n = 1,198), Leipzig (5.92 %; n = 711), Moers (3.19 %; n = 711), and Bonn (3.19 %; n = 318). One hundred and eight administrations (5.01 %) published fewer than 20 records. Thus, the range of published Open Data is between one dataset (min.) and 2,198 (max.), with a mean of 136 datasets. The majority of datasets (50.89 %; n = 6,117) belong to the Spatial Planning theme. The Spatial Planning theme is divided by development plans (n = 5,485), urban areas (n = 135), districts (n = 57), real estate cadaster (n = 40), addresses (n = 17), orthophotos (n = 8), blocks (n = 3), land uses (n = 3), house numbers (n = 3), postal code areas (n = 2), and social areas (n = 2). The second largest data set, population, accounts for only 10.87% (n = 1,306) of the data. This category is split by designation such as demographics (n = 501), population (n = 233), migration background (n = 150), first names (n = 103), employment (n = 90), citizenship (n = 74), housing (n = 71), births and deaths (n = 35), communities in need (n = 32), religious affiliation (n = 11), people with disabilities (n = 3), integration (n = 2), and refugee numbers (n = 1).

Examination of published data concerning location factors

Of the 12,020 published data sets, more than half of the data (57.50 %; n = 6,912) are classified as hard location factors. However, most hard location factors are represented by land-use planning data in development plans (48.63 %; n = 5,845). Important hard location factors such as office, industrial and commercial space, information on business development, and business locations are missing or only available to a minimal extent. Other important infrastructural information, such as heavy-load traffic, highway access, road traffic, bridges, airports, and shipping docks for potential decision-makers, is also missing or is only available in isolated cases. Thus, it can be stated that the voids of Open Data regarding hard location factors are multifaceted. The soft business-related location factors only account for a small share (7.15 %; n = 859). Here, data on demographics are primarily published (4.17 %; n = 501), which is relevant for potential location decision-makers but is far from a unique selling point. For example, there is a lack of data on student numbers, location development, crime statistics, childcare places for daycare facilities, and many other categories. Thus, it can be stated that significant data gaps appear in the context of Open Data for the soft business-related location factors. Furthermore, the soft person-related location factors are to be examined. These soft person-related location factors also account for only a small part of Open Data (4.32 %; n = 519). The person-related location factors show a greater diversity of data than the hard and business-related location factors. For example, although there is a data focus on the publication of student numbers (0.68 %; n = 82), other person-related location factors are also published in the form of Open Data. These include Open Data in the areas of hospitals (n = 41), bicycle routes (n = 36), facilities in the area of sports and playgrounds (n = 33), green spaces or parks (n = 32), and weekly markets (n = 32). This diversity of data can be explained by the fact that the personal location factors also overlap in content with the civil society target group of Open Data. Almost one third of the published Open Data does not relate to the location decision (31.03 %; n = 3,730). This data includes landmarks, campgrounds, tree cutting, historical aerial photographs, honorary graves, and more. This variety is initially plausible since Open Data has multiple addressees besides the business community (civil society, media, research). However, it is already apparent that data is only inadequately supplied to the business community as a stakeholder group. This fact leads to the assumption that this is also the case in other stakeholder groups.

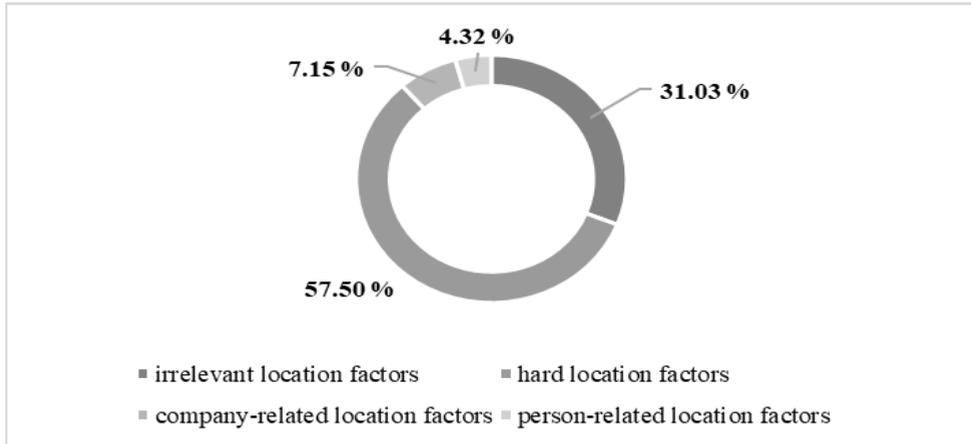


Figure 1. Examination of published data concerning location factors.

Source: own processing according to Bertelsmann Stiftung, 2021 (n = 12,020)

In addition to the conspicuous data gaps concerning location factors, another finding is emerging: Not all actors publish open data in all categories of the 176 cities, local communities, and municipalities in the sample data catalog. The ten largest data focal points are by no means filled with data by all players. While the share in the category of regional planning is still 78.98 % (share of actors 139 of 176), this shrinks dramatically in other location-relevant categories. Less than one third of the actors in the location-relevant categories such as population (31.82 %), infrastructure (31.82 %), private transport (27.27 %), construction (21.59 %), schools (19.32 %), and energy industry (7.39 %) publish Open Data.

Table 1. Top ten categories of open data.

Rang	Category	n =	Share of stakeholders	Share of stakeholders in %
1	Spatial planning	6.117	139 of 176	78.98 %
2	Population	1.306	56 of 176	31.82 %
3	Elections	875	82 of 176	46.59 %
4	Infrastructure	373	56 of 176	31.82 %
5	Individual transport	319	48 of 176	27.27 %
6	Energy industry	290	13 of 176	7.39 %
7	Construction	217	38 of 176	21.59 %
8	Schools	146	34 of 176	19.32 %
9	Budget	143	23 of 176	13.07 %
10	Website	132	4 of 176	2.27 %

Source: own processing according to Bertelsmann Stiftung, 2021 (n = 9,918)

5 Conclusion

Open Data can compare data systematically. In doing so, this data can be used for a variety of purposes. One way of using Open Data is in location analysis. However, this requires a uniform data structure. The sample data catalog offers this uniformity for the first time. Furthermore, it forms the basis for answering the research question: Which Open Data (location factors) are published on the part of the municipalities and which Open Data (location factors) are not published so far in order to increase the own location attractiveness?

First of all, it can be stated that many relevant location factors such as development plans (hard location factor), data on the demographics of a region (business-related location factors) as well as data on hospitals, bike routes, and facilities in the area of sports and play venues (person-related location factors) are already published. Particularly in the category of person-related location factors, a range of data can be seen in terms of content. This range is because Open Data also addresses other target groups besides the economy, such as the civil society. The overlap can therefore explain the diversity. However, many data gaps remain open that are relevant for location analyses. These data gaps are significant and are not only of content or qualitative nature. For example, not all 176 cities, local municipalities, and association municipalities publish Open Data in all categories. Significant quantitative data gaps are evident here. These data gaps have an impact on the attractiveness of the location. If data are not available to potential location decision-makers, they cannot be compared. As a result, the visibility of a location decreases, and the location attractiveness is weakened. It is therefore crucial for cities, local communities, and municipalities to improve this. If open data are not published, the attractiveness of a location in the region will drop considerably.

References

1. Begany, G., & Gil-Garcia, R. (2021). Understanding the actual use of open data: Levels of engagement and how they are related. *Telematics and Informatics*, *63*, 1–16.
2. Bertelsmann Stiftung. (2021). *Musterdatenkatalog für Kommunen - Welche offenen Daten werden von Kommunen in Deutschland veröffentlicht?* <https://www.bertelsmann-stiftung.de/de/unsere-projekte/smart-country/musterdatenkatalog>.
3. Conway, E. (2011). Protektionismus. In E. Conway (Ed.), *50 Schlüsselideen Wirtschaftswissenschaft* (pp. 170–173). Spektrum Akademischer Verlag.
4. Dapp, M., Balta, D., Palmethofer, W., Krcmar, H., & Kuzev, P. (2016). *Open Data. The Benefits: Das volkswirtschaftliche Potential für Deutschland*. Konrad-Adenauer-Stiftung e.V.
5. Etezadzadeh, C. (2020). *Smart City – Made in Germany*. Springer Fachmedien Wiesbaden.
6. European Commission. (2020). *Was sind offene Daten?* Europeandataportal. <https://www.europeandataportal.eu/elearning/de/module1/#/id/co-01>.
7. Godau, M. (2001). *Die Bedeutung weicher Standortfaktoren bei Auslandsinvestitionen mit besonderer Berücksichtigung des Fallbeispiels Thailand* (RWTH-CONV-118637) [Aachen]. RIS. <https://publications.rwth-aachen.de/record/56543>.
8. Klenk, T., Nullmeier, F., & Wewer, G. (2020). *Handbuch Digitalisierung in Staat und Verwaltung*. Springer Fachmedien Wiesbaden.
9. König, C. (2021). *Wie ein Grossprojekt eine ganze Region verändert*. Kommunal, *9*.
10. Khurshid, M., Zakaria, N., Rashid, A., Kazmi, R., Shafique, M., & Ahmad, M. (2019). Analyzing diffusion patterns of big open data as policy innovation in public sector. *Computers and Electrical Engineering*, *78*, 1–14.
11. Lassinantti, J., Ståhlbröst, A., & Runardotter, M. (2019). Relevant social groups for open data use and engagement. *Government Information Quarterly*, *36*, 1–14.
12. Lennardt, S., & Stakemeier, D. (2017). *Standortmarketing in der Wirtschaftsförderung: Grundlagen für die Praxis. Wirtschaftsförderung in Lehre und Praxis*. Springer Fachmedien Wiesbaden.

13. Leviakangas, P., & Molarius, R. (2020). Open government data policy and value added - Evidence on transport safety agency case. *Technology in Society*, *63*, 1–9.
14. Liao, Y., Zhou, Q., & Jing, X. (2021). A comparison of global and regional open datasets for urban greenspace mapping. *Urban Forestry & Urban Greening*, *62*, 1–12.
15. Mergela, I., Kleibrink, A., & Sörvik, J. (2018). Open data outcomes: U.S. cities between product and process innovation. *Government Information Quarterly*, *35*, 622–632.
16. Meyer-Stamer, J. (2013). *Lokale und regionale Standortpolitik - Konzepte und Instrumente jenseits von Industriepolitik und traditioneller Wirtschaftsförderung* (<http://inef.uni-due.de/cms/files/report39.pdf>). SUB Hamburg Carl von Ossietzky. <http://edoc.vifapol.de/opus/volltexte/2013/4567>.
17. Nikiforova, A., & McBride, K. (2021). Open government data portal usability: A user-centred usability analysis of 41 open government data portals. *Telematics and Informatics*, *58*, 1–13.
18. Preische, J. (2014). *Digitales Gold Nutzen und Wertschöpfung durch Open Data für Berlin*. TSB Technologiestiftung Berlin.
19. Runeson, P., Olsson, T., & Linåker, J. (2021). Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept. *Journal of Systems and Software*, *182*, 1–17.
20. Statista. (2021a). *Gemeinden in Deutschland* (did-102105-1).
21. Statista. (2021b). *Unternehmenslandschaft in Deutschland* (did-23690-1).
22. von Lucke, J. (2010). *Open Government - Öffnung von Staat und Verwaltung; Deutsche Telekom Institute for Connected Cities: Friedrichshafen*.
23. Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data – A critique of open government datasets in the UK. *Government Information Quarterly*, *37*, 1–10.