Senatsverwaltung
für Stadtentwicklung
und Wohnen

Berlin

Big Data und Crowd Data für die Berliner Stadtentwicklungsplanung

Zürich / Berlin
27.04.2017

Auftraggeberin

SenStadtWohn
Referat Stadtentwicklungsplanung

Dr. Paul Hebes (Projektleitung)
Tel. +49 (0)30 9025 1326
paul.hebes@sensw.berlin.de

Elke Plate
Tel. +49 (0)30 9025 1328
Elke.plate@sensw.berlin.de

Thorsten Tonndorf
Tel. +49 (0)30 9025 1327
Thorsten.Tonndorf@sensw.berlin.de

Projektbearbeitung

EBP Schweiz AG
Mühlebachstrasse 11
8032 Zürich
Tel. +41 44 395 16 16
info@ebp.ch
www.ebp.ch

EBP Deutschland GmbH
Am Hamburger Bahnhof 4
10557 Berlin
Germany
Tel. +49 30 120 86 82-0
info@ebp.de
http://www.ebp.de

Zusammenfassung

Das Wachstum Berlins gestalten

Die Senatsverwaltung für Stadtentwicklung und Wohnen steht vor der Herausforderung, das starke Wachstum Berlins mit anhaltenden wirtschaftlichen und gesellschaftlichen Transformationsprozessen erfolgreich, nachhaltig und sozialverträglich zu gestalten. Die strategische Planung aktualisiert in diesem Zusammenhang die Stadtentwicklungspläne Wohnen, Zentren sowie Industrie und Gewerbe.

Potenziale von Big Data und Crowd Data für die Stadtentwicklungsplanung prüfen

Ziel des Referats Stadtentwicklungsplanung ist es, für die Arbeiten an den Stadtentwicklungsplänen vorhandene und neue Daten und Erkenntnisse zu nutzen. In diesem Sinne zielt die Arbeit darauf, durch den Nutzen von Big Data und Crowd Data Innovationen für die strategische Planung zu generieren. Es ist dabei auch zu prüfen, inwieweit Big Data und Crowd Data für die Stadtentwicklungsplanung mit den Herausforderungen der wachsenden Stadt künftig stärker Verwendung finden können. Dazu gilt es herauszuarbeiten, welche Potenziale die Daten für die Stadtentwicklung bieten und wie sie künftig neu oder stärker genutzt werden können, um Berlin nachhaltig, sozial und wirtschaftlich erfolgreich zu gestalten.

Zentrale Begriffe

„Big Data" ist definiert als große, vielfältige, dynamische und/oder unstrukturierte private oder öffentliche Datenströme mit unklarer Vertrauenswürdigkeit. Innerhalb von Big Data können die drei Typen der nutzergenerierte, transaktionsgenerierten und sensorgenerierten Daten unterschieden werden. „Crowd Data" bezeichnet Daten, die von einer meistens sehr großen Gruppe von Personen oftmals dezentral und nach relativ informellen Regeln erarbeitet werden. Dies wird als „Crowdsourcing" bezeichnet.

Bei der Betrachtung der generellen Nutzungspotenziale für die Stadtentwicklungsplanung sind die drei Typen von Big Data sowie Crowd Data spezifisch zu betrachten.

Nutzergenerierte Daten: Potenzial mit erschwerter Nutzung

Bei nutzergenerierten Daten lassen sich aus der Überlagerung von individuellen Nutzungs- bzw. Verhaltensmustern im Raum Aussagen zur Anzahl bzw. zur Dichte sowie der Mischung der Nutzer*innen an spezifischen Orten gewinnen. Da nutzergenerierte Daten von Personen durch ihr Handeln im Raum generiert werden, vermitteln sie den Eindruck, dass damit „reale" Verhaltensweisen und Präferenzen ablesbar sind. Die Informationen erfordern jedoch eine Interpretation von Motivlagen, was die Gefahr von Fehl- oder Überinterpretation mit sich bringt. Zudem bestehen verschiedene Hürden, die eine Nutzung der Daten erschweren: eine eingeschränkte Zugänglichkeit, eine häufig aufwändige datentechnische und analytische Aufbereitung und eine vielfach geringe und häu-

fig kaum beurteilbare Datenqualität. Hinzu kommt die Schwierigkeit, angesichts der hohen Volatilität der Plattformen Vergleiche über mittel- und langfristige Zeiträume erstellen zu können.

Transaktionsgenerierte Daten: Nutzung abhängig vom Geschäftsmodell der Anbieter

Transaktionsgenerierte Daten weisen unterschiedliche Nutzungspotenziale auf.

- Bei Daten aus webbasierten Transaktionen sind, bei vergleichbaren Nutzungsmöglichkeiten, Zuverlässigkeit, Genauigkeit und Aussagekraft höher einzuschätzen als bei den nutzergenerierten Daten.
- Die generelle Verfügbarkeit privater Unternehmensdaten dürfte kaum gegeben sein. Für die Stadtentwicklungsplanung ist jedoch im Austausch mit den Unternehmen eine Nutzung in konkreten Planungen denkbar.
- Die konkreten Nutzungsmöglichkeiten von Daten öffentlicher Unternehmen wären spezifisch abzuklären.

Sensorgenerierte Daten: Großes Potenzial mit aufwändiger Nutzung

Über Sensoren wird eine enorme Menge an Daten gesammelt. Sensorgerierte Daten sind auch dadurch charakterisiert, dass sie oft in (Nah-)Echtzeit zur Verfügung stehen. Mit dem Internet der Dinge bzw. Internet of Things (IoT) dürfte sich die Datenmenge und damit das Nutzungspotenzial weiter erhöhen. Die Analyse und Nutzung von Echtzeit-Sensordaten erfordert jedoch umfassende Kompetenzen und Ressourcen. Bei sensorgenerierten Daten ist eine subjektive Verzerrung des Gemessenen gering. Sie sind damit vergleichsweise valide. Sensorgenerierte Daten sind oft auch zentrale Elemente von Smart City-Konzepten und werden bereits heute zur Regelung und Steuerung eingesetzt, etwa im Energie- oder Verkehrsbereich.

Crowd Data: Potenzial als zusätzlicher Kanal der Bürgerkommunikation

Crowdsourcing bietet eine neue bzw. ergänzende Möglichkeit einer gezielten und unmittelbaren Interaktion mit der Stadtbevölkerung. Der steigende Grad der digitalen Vernetzung, die ständige Weiterentwicklung der technischen Möglichkeiten und die zunehmend einfacher werdende Handhabung von mobilen Angeboten werden die Möglichkeiten des Crowdsourcings weiter ausdehnen.

Nutzungspotenzial für die Stadtentwicklungspläne

Ratingplattformen, Immobilien-, Unterkunfts- und Mobilitätsportale besitzen die höchste Aussagekraft und Relevanz für die drei betrachteten StEPs. Damit lassen sich etwa Aussagen zur Ausstattung oder Attraktivität von Zentren, zum Nachfrageverhalten, zu Einzugsbereichen oder zur Umfeld- oder Standortqualität treffen. Die Daten können dabei einen Eindruck eines Sachverhaltes vermitteln und Hinweise liefern auf näher zu betrachtende Zusammenhänge oder Fragestellungen. Dies insbesondere in Kombination bzw. durch die Überlagerung mit weiteren Daten.

Weitere Sondierung notwendig

Die skizzierten Anwendungsoptionen in den einzelnen Stadtentwicklungsplänen lassen noch keine Aussagen zu konkreten Nutzungsmöglichkeiten zu. Es wäre weiter zu sondieren, ob und wenn ja, welche Analysen im Zusammenhang mit den einzelnen Stadtentwicklungsplänen durch die Auswertung von Big Data und Crowd Data untersetzt werden könnten. Dies mit dem Ziel, weitergehenden Planungsansätzen zusätzlich erforderliche Datengrundlagen zur Verfügung zu stellen.

Einschätzungen zu Chancen und Risiken im Umgang mit Big Data und Crowd Data

Die folgende Übersicht zeigt eine zusammenfassende Einschätzung der Chancen und Risiken im Umgang mit Big Data und Crowd Data aus Sicht der Stadtentwicklungsplanung.

Chancen	**Risiken**
Verfügbarkeit neuer Daten, lassen neuartige Zusammenhänge erkennen und Verbindungen herstellen	Nutzungsmöglichkeiten (kurzfristig) überschätzt
Räumlich hoch aufgelöste und feingliedrige Daten, ermöglichen Rückschlüssen auf kleinräumiger Ebene	Sammlung, Verarbeitung und Auswertung der Daten stellt erhöhte Anforderungen an Ressourcen und Kompetenzen
Zeitnahe Verfügbarkeit, ermöglicht ein schnelleres Erkennen von Entwicklungen und eine verkürzte Reaktionszeit	Falsche Interpretationen oder Trugschlüsse aufgrund von Momentaufnahmen
Geringerer Aufwand für die Datenerhebung und -beschaffung	Teilweise eingeschränkte Datenqualität, Volatilität erschwert Zeitreihen
Reduktion der non-response-Rate klassischer Erhebungs- und Befragungsinstrumente	Steigende Datenmenge erhöht die Komplexität und kann Entscheidungen verzögern oder erschweren
Neue und bessere Daten unterstützen bedarfsgerechtere Entscheidungen und Planungen	Nutzung von Big Data wirft ethische Fragen auf bezüglich Datenschutz und Privatsphäre
Crowdsourcing als Möglichkeit einer partizipativeren, offeneren und bürger*innenorientierten Stadtentwicklung	Umgang mit nicht-hoheitlichen Daten in hoheitlichen Planungen und Prozessen wirft rechtliche Fragen auf
	Datenschutzrechtliche Bestimmungen erschweren explorative Datennutzung

Vertiefte Auseinandersetzung mit Big Dat und Crowd Data

Es ist davon auszugehen, dass sich Big Data rasant weiter entwickeln wird, mit wachsenden Datenmengen sowie verbesserten Analysemethoden und Know-how in der Datenauswertung. Entsprechend erscheint eine vertiefte Auseinandersetzung mit Big Data und Crowd Data lohnenswert. Die Betrachtung der Nutzungsmöglichkeiten in der Stadtentwicklungsplanung steht dabei im Kontext eines größeren gesellschaftlichen Diskurses um Datennutzung, Datenschutz und Privatsphäre.

Inhaltsverzeichnis

1 Ausgangslage und Ziele ... 1
1.1 Ausgangslage ... 1
1.2 Ziele ... 1
2 Projektbeschreibung ... 3
2.1 Grundverständnis ... 3
2.2 Aufbau und Vorgehen ... 4
2.2.1 Aufbau des Projektes ... 4
2.2.2 Vorgehen ... 4
3 Begriffliche Grundlagen ... 6
3.1 Big Data ... 6
3.1.1 Begriffsdefinition ... 6
3.1.2 Daten, Information, Wissen ... 8
3.1.3 Typisierung ... 9
3.1.4 Abgrenzung: Big Data und Small Data ... 9
3.2 Crowd Data ... 10
4 Datenquellen ... 12
4.1 Kategorien von Datenquellen ... 12
4.1.1 Beschreibung der Datenquellen ... 12
4.1.2 Möglichkeiten stadtraumbezogener Aussagen ... 14
4.2 Einschätzung zu den Datenquellen ... 17
4.2.1 Georeferenzierung ... 17
4.2.2 Zeitreihen ... 19
5 Bezugskanäle ... 22
5.1 Bezugskanäle im Überblick ... 22
5.2 Betrachtung der einzelnen Bezugskanäle ... 23
5.2.1 APIs ... 23
5.2.2 Webscraping ... 24
5.2.3 Einkauf bei der Datenquelle ... 25
5.2.4 Datenbroker ... 25
5.2.5 Community ... 26
5.3 Einschätzung zu den Bezugskanälen ... 27
5.3.1 Zugänglichkeit ... 27
5.3.2 Zeitreihen ... 28
5.3.3 Datenqualität ... 28
6 Nutzungspotenzial für die Stadtentwicklungsplanung ... 30
6.1 Nutzungspotenzial von nutzergenerierten Daten ... 30
6.2 Nutzungspotenzial transaktionsgenerierter Daten ... 34

6.3 Nutzungspotenzial sensorgenerierter Daten 36
6.4 Nutzungspotenzial von Crowd Data 38

7 Nutzungspotenzial für die Stadtentwicklungspläne (StEP) 41
7.1 StEP Zentren 3 41
7.1.1 Beschreibung 41
7.1.2 Datenquellen mit Aussagekraft zumStEP Zentren 42
7.2 StEP Wohnen 44
7.2.1 Beschreibung 44
7.2.2 Datenquellen mit Aussagekraft zum StEP Wohnen 44
7.3 StEP Industrie und Gewerbe 46
7.3.1 Beschreibung 46
7.3.2 Datenquellen mit Aussagekraft zum StEP Industrie und Gewerbe 46

8 Zusammenfassende Einschätzungen und Ausblick 48
8.1 Chancen und Risiken im Umgang mit Big Data und Crowd Data – Erkenntnisse des Fachgesprächs 48
8.1.1 Chancen im Umgang mit Big Data und Crowd Data 49
8.1.2 Risiken im Umgang mit Big Data und Crowd Data 50
8.2 Ansatzpunkte zur vertieften Auseinandersetzung mit Big Data und Crowd Data 53
8.2.1 Eigene Kompetenzen aufbauen 53
8.2.2 Den verwaltungsübergreifender Austausch verstärken 54
8.2.3 Externe Netzwerke etablieren 54
8.2.4 Konkrete Nutzungsmöglichkeiten prüfen 55
8.2.5 Erfahrungen mit Crowdsourcing sammeln 56

Anhänge

A1 Literatur

A2 Übersicht Anwendungsbeispiele

A3 Beispiele für Crowdsourcing

A4 Beispiele für Small Data

A5 Sammlung Datenquellen

A6 Beschreibung ausgewählter Bezugskanäle

A7 Fachgespräch

1 Ausgangslage und Ziele

1.1 Ausgangslage

Daten als „Öl des 21. Jahrhunderts“

Im Zeitalter der Digitalisierung scheinen Daten überall gegenwärtig und unendlich verfügbar zu sein: Digitalisierung produziert Daten, das „Öl des 21. Jahrhunderts”, in ungeahntem Ausmaß. Welche Bedeutung hat Big Data für Gesellschaft, Politik und Verwaltung?

Große Potenziale, aber mit unklaren Grenzen

Die Einschätzung ist ambivalent: Daten erscheinen zugleich über- und unterbewertet zu sein. Sie sind unterbewertet, da von datengenerierenden Systemen scheinbar endlose Mengen von Bits und Bytes produziert werden. „Gräbt“ man ausreichend tief in diesen Daten, so die Vermutung, lassen sich ganz neue Zusammenhänge erkennen und Wissen generieren. Die Daten sind zugleich potenziell überbewertet, weil alle Welt von der „Macht der Daten“ spricht, ohne sich im Klaren darüber zu sein, welches die Grenzen des neu zu gewinnenden Wissens sind.

Big Data verspricht Effizienzsteigerungen

Entwicklungen in Sensorik, Informationstransfer und -verarbeitung ermöglichen es, in städtischen Räumen Umwelt- und Infrastrukturdaten in Quasi-Echtzeit zu erfassen. So verspricht Big Data Effizienzsteigerungen für Städte und Gemeinden, aber auch für Unternehmen und Privatpersonen. Allerdings bedingt dies die weniger diskutierte Möglichkeit, nach einer Datenauswertung Informationen und Verhaltensanreize an die Akteure zurückzuspielen.

Vieles in Bewegung, aber noch Weniges greifbar

Auf einen Nenner gebracht: Big Data hat mit rasanter Geschwindigkeit vieles in Bewegung gesetzt und Erwartungen geweckt, Art und Ausmaß konkreter Veränderungen sind jedoch noch nicht abzusehen. Auch die Möglichkeiten, Potenziale und Herausforderungen einer künftigen Nutzung von Big Data in der Berliner Stadtentwicklungsplanung können bisher kaum abgeschätzt werden.

1.2 Ziele

Das Wachstum Berlins gestalten

Die Senatsverwaltung für Stadtentwicklung und Wohnen steht vor der Herausforderung, das starke Wachstum Berlins mit anhaltenden wirtschaftlichen und gesellschaftlichen Transformationsprozessen erfolgreich, nachhaltig und sozialverträglich zu gestalten. Die strategische Planung aktualisiert in diesem Zusammenhang die Stadtentwicklungspläne Wohnen, Zentren sowie Industrie und Gewerbe.

Nutzbare Potenziale erkennen

Ziel des Referats Stadtentwicklungsplanung ist es, für die Arbeiten an den Stadtentwicklungsplänen vorhandene und neue Daten und Erkenntnisse zu nutzen. In diesem Sinne zielt die Arbeit darauf, durch den Nut-

zen von Big Data und Crowd Data Innovationen für die strategische Planung zu generieren. Es ist dabei auch zu prüfen, inwieweit Big Data und Crowd Data für die Stadtentwicklungsplanung mit den Herausforderungen der wachsenden Stadt künftig stärker Verwendung finden können. Dazu gilt es herauszuarbeiten, welche Potenziale die Daten für die Stadtentwicklung bieten und wie sie künftig neu oder stärker genutzt werden können, um Berlin nachhaltig, sozial und wirtschaftlich erfolgreich zu gestalten.

Strukturierter Überblick und systematische Erfassung

Das Projekt legt dazu die Grundlagen, indem es:

- einen strukturierten Überblick über die Thematik schafft,
- eine systematische Erfassung und Beurteilung von Big Data-Quellen und Bezugskanälen vornimmt,
- Nutzungspotenziale von Big Data und Crowd Data für die Stadtentwicklungsplanung in ihren vielfältigen Aufgaben in den Themen Zentren, Wirtschaft und Wohnen abschätzt,
- in einem ersten Fachgespräch mit Expert*innen unterschiedliche Sichtweisen und Einschätzungen zu Nutzungspotenzialen und Chancen sowie Risiken im Umgang mit Big Data und Crowd Data erörtert,
- skizziert, wie sich die Stadtentwicklungsplanung künftig vertieft mit Big Data und Crowd Data auseinandersetzen und ihre Erfahrungen und Kompetenzen vertiefen kann.

2 Projektbeschreibung

Kapitel 2 legt die grundsätzliche Zugangsweise des Projektes dar und beschreibt die Struktur des Projektes sowie das Vorgehen in der Bearbeitung.

2.1 Grundverständnis

Die Vorgehensweise basiert auf folgendem Grundverständnis:

Explorativer Zugang

Angesichts der Neuartigkeit des Themas existiert kein klar erkennbarer „Wissenskanon" zu Big und Crowd Data, insbesondere nicht aus Sicht der (Stadt-)Planung. Das Projekt hat deshalb einen explorativen Charakter. Es will die Thematik erkunden und strukturieren. Dabei sind die inhaltliche Breite und die Übersicht bedeutsamer als die Bearbeitungstiefe in einzelnen Punkten.

Thematisch breit und offen

Der Zugang erfolgt über den Begriff „Big Data". Damit soll die Thematik möglichst breit und offen betrachtet werden, um sie anschließend auf die Stadtentwicklungsplanung beziehen zu können.

Fach- und Technikperspektive

Die Bearbeitung erfolgt im Zusammenspiel von themenbezogenen Kenntnissen der Stadtentwicklungsplanung sowie der Kompetenz im Umgang mit räumlichen Daten. Dies stellt sicher, mit dem Projekt zu fachlich sinnvollen und technisch machbaren Anwendungen zu gelangen.

Grundlage für Weiterführung

Angesichts der thematischen Breite können die Ergebnisse nicht abschließend sein. Das Projekt legt jedoch die Basis dafür, die Thematik in Berlin weiterzudenken und Themen und Fragen für weiterführende Aktivitäten zu identifizieren.

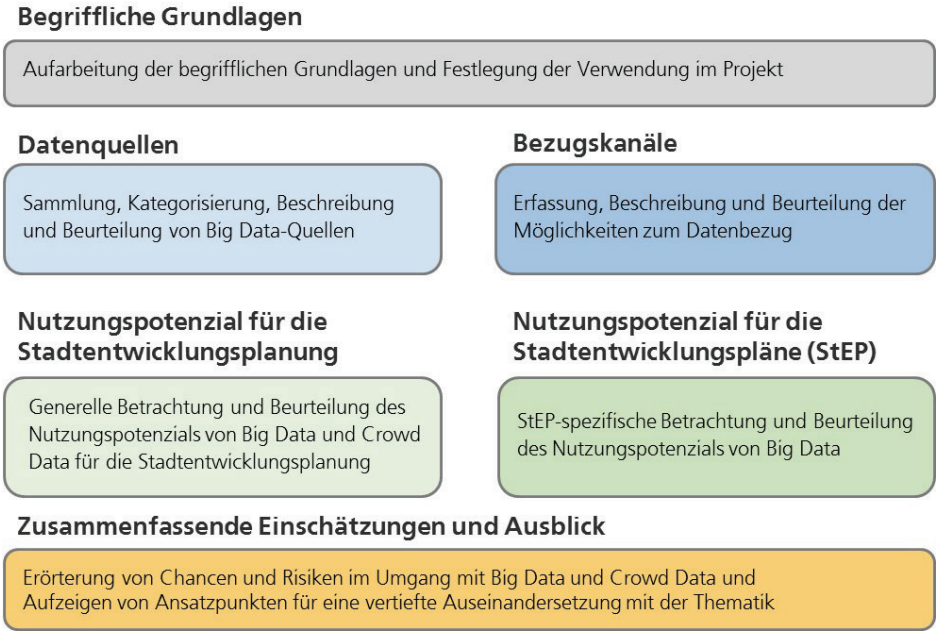
2.2 Aufbau und Vorgehen

2.2.1 Aufbau des Projektes

Sechs Elemente

Das Projekt gliedert sich in sechs Elemente (Abbildung 1). Sie geben gleichzeitig die Struktur dieses Berichtes vor.

Abbildung 1: Aufbau des Projekts



2.2.2 Vorgehen

Im Folgenden werden die Schritte für die Erarbeitung der sechs Berichtsbestandteile dargelegt.

Begriffliche Grundlagen

Im ersten Schritt wurden die Begriffe aufgearbeitet und ihre Verwendung im Projekt festgelegt. Dies stützt sich auf eine Literaturrecherche zu den Begriffen „Big Data“ und „Crowd Data“.

Datenquellen

Mit einer Literaturanalyse von Anwendungsbeispielen, in denen Big Data-Quellen für eine räumliche Analyse genutzt wurden, und im Rahmen einer Internetrecherche wurden mögliche Datenquellen gesammelt. Die Datenquellen wurden zuvor identifizierten Typen von Big Data zugeordnet, strukturiert und anschließend kategorisiert. Die entsprechende Übersicht findet sich in Anhang A5. Jede der Kategorien von Datenquellen wurde anschließend dahingehend beschrieben, welche stadtraumbezogenen Aussagen sich grundsätzlich ableiten lassen. Zudem werden die Datenquellen im Hinblick auf eine mögliche Nutzung beurteilt.

Bezugskanäle

Zur Erfassung von Bezugskanälen von Daten wurde eine entsprechende Recherche durchgeführt. Ausgewählte Bezugskanäle wurden soweit beurteilbar hinsichtlich ihrer Zugänglichkeit, ihrer Qualität sowie ihrer räumlichen, zeitlichen und inhaltlichen Auflösung beschrieben (Anhang

A6). Auch bei den Datenquellen wurde eine Einschätzung hinsichtlich einer möglichen Nutzung vorgenommen.

Generelles Nutzungspotenzial für Stadtentwicklungsplanung

Die Einschätzung zum generellen Nutzungspotenzial basiert auf einer Literaturrecherche, den Fachkenntnissen der Autoren, inhaltlichen Diskussionen mit der Auftraggeberin sowie Inputs aus einem Fachgespräch in einem erweiterten Kreis (vgl. Anhang A7). Die Betrachtung erfolgt differenziert nach den unterschiedlichen Typen von Big Data bzw. Crowd Data.

Spezifischer Nutzen für die STEPs

Ausgehend von dem Anliegen, die Stadtentwicklungspläne innovativ auszugestalten und neue Erkenntnisquellen durch Big Data und Crowd Data zu erschließen, werden Big Data Quellen und deren Aussagekraft zu den Themen der drei Stadtentwicklungspläne (StEP) „Zentren", „Wohnen" sowie „Industrie und Gewerbe" aufgezeigt. Diese drei Stadtentwicklungspläne wurden ausgewählt, da sie sich in Überprüfung oder Fortschreibung befinden. Die Resultate des Projekts können somit einen Beitrag zu deren Aktualisierung liefern.

Zusammenfassende Einschätzungen und Ausblick

Maßgeblich gestützt auf die Diskussionen in einem Fachgespräch (Anhang A7) werden in Kapitel 8 Chancen und Risiken im Umgang mit Big Data und Crowd Data erörtert. Basierend auf den verschiedenen Analyseschritten werden letztlich unterschiedliche Ansatzpunkte und konkrete Maßnahmen aufgezeigt, wie sich die Stadtentwicklungsplanung vertieft mit Big Data und Crowd Data auseinandersetzen und verstärkt für ihre Zwecke nutzbar machen kann.

3 Begriffliche Grundlagen

In Kapitel 3 werden die beiden zentralen Begriffe „Big Data“ und „Crowd Data“ aufgearbeitet und ihre Verwendung im Projekt festgelegt.

„Big Data“ ist definiert als große, vielfältige, dynamische und / oder unstrukturierte private oder öffentliche Datenströme mit unklarer Vertrauenswürdigkeit. Innerhalb von Big Data können die drei Typen der nutzergenerierte, transaktionsgenerierten und sensorgenerierten Daten unterschieden werden. „Crowd Data“ bezeichnet Daten, die von einer meistens sehr großen Gruppe von Personen oftmals dezentral und nach relativ informellen Regeln erarbeitet werden (als „Crowdsourcing“ bezeichnet).

3.1 Big Data

3.1.1 Begriffsdefinition

Keine allgemeingültige Definition

Der Begriff „Big Data“ wird in vielfältigen Zusammenhängen mit unterschiedlichen Definitionen und Beschreibungen genutzt. Eine allgemeingültige Definition existiert nicht[1].

Charakterisierung durch vier Vs

In der Literatur werden zur Charakterisierung von Big Data aber häufig vier Vs verwendet (etwa BITKOM 2014). In dieser Lesart zeichnen sich Big Data durch die spezielle Ausprägung von mindestens einem der Vs oder von einer Kombination von Vs aus:

- **Volume**: Große Datenmengen
- **Variety**: Vielfalt an unterschiedlichen Datenquellen, -formaten und -modellen und unstrukturierte Daten
- **Velocity**: Hohe Geschwindigkeit der Datengenerierung und -auswertung, nicht selten in Echtzeit
- **Veracity**: A priori unsichere Wahrhaftigkeit von Daten

Teilweise werden noch weitere Eigenschaften ergänzt, wie etwa „Value“.

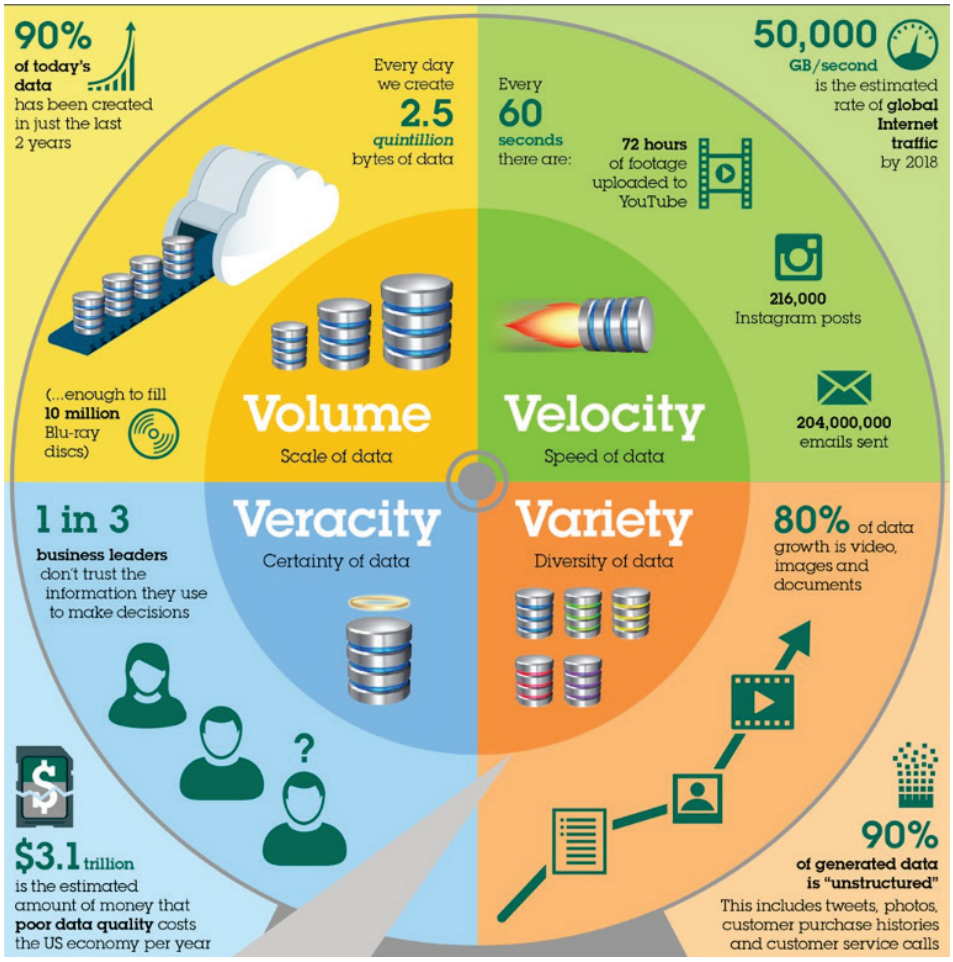
Was ist neu an Big Data?

Die Berner Fachhochschule (BFH 2015:6) hält in einer Studie für die Bundesregierung der Schweiz fest, dass keine dieser Dimensionen und auch keine Kombination davon grundsätzlich neu sind. „Neu an Big Data ist, dass die Nutzung von Daten in allen Bereichen stark zunimmt, dass die Auswertung systematisch durchgeführt wird, und dass der Datenzusammenzug massiv über bestehende Domänengrenzen hinausgeht, wie dies

[1] Ward, Baker (2013)

beispielsweise beim Zuwachs an verfügbaren Datenquellen im Internet of Things (IoT) oder in der Kombination verschiedener personenbezogener sensorgenerierter Daten im Social Mobile Computing zu beobachten ist." (ibid.)

Abbildung 2: The Four Vs of Big Data (IBM 2016)



Verwendung im Projekt

Im vorliegenden Projekt wird Big Data definiert als „große, sehr vielfältige, sich dynamisch verändernde und oft unstrukturierte private und öffentliche Datenströme". Weitere Definitionen schließen etwa die Datenanalyse bzw. deren Nutzung ebenfalls ein, beispielsweise BITKOM (2014)[2]. Im Rahmen des vorliegenden Projekts wird aber die engere Definition genutzt, welche sich auf die eigentlichen Daten beschränkt.

Verarbeitbarkeit als konstituierendes Element

Die Komplexität der Datenverarbeitung ist ein wichtiges konstituierendes Element, welches auch in umgangssprachlichen Definitionen von Big Data seinen Ausdruck findet, wie „nicht mehr auf einem herkömmlichen

[2] „Big Data [ist der] Einsatz großer Datenmengen aus vielfältigen Quellen mit einer hohen Verarbeitungsgeschwindigkeit zur Erzeugung wirtschaftlichen Nutzens"

Rechner verarbeitbare Daten“ oder „any data that cannot fit into an Excel spreadsheet“ (Batty 2013).

3.1.2 Daten, Information, Wissen

Von Daten zu Informationen

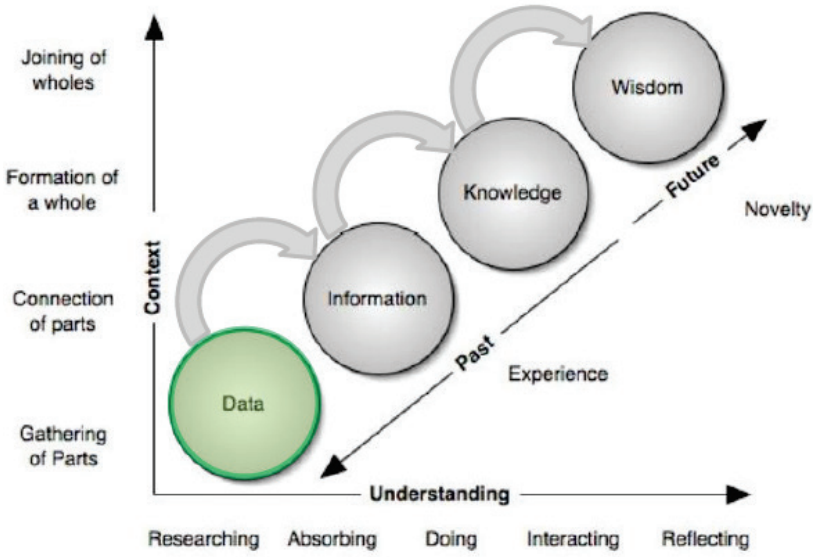
Ein Trugschluss ist zu vermeiden: Die immer umfangreicher verfügbaren Daten dürfen nicht mit Information oder gar Wissen gleichgesetzt werden. Bei Daten handelt es sich um Symbole und Zeichen, die Signale repräsentieren. Sie bergen keine inhärente Bedeutung oder Nützlichkeit. Durch Umformung, Strukturierung (Mustererkennung), „Destillation“ und Interpretation von Daten durch eine analysierende Person entsteht aus Daten Bedeutung, also Information.

Von Informationen zu Wissen

Durch Wahrnehmung, Erfahrung, Werte, Kontextinformationen, Experteneinsicht und Intuition geformte, organisierte oder im richtigen Kontext erinnerte Information bezeichnen wir als Wissen, welches etwa für Strategieprozesse oder zur Entscheidungsfindung in der Stadtentwicklungsplanung genutzt werden kann. Wissen ist – im Gegensatz zu Daten und Informationen – a priori persönlich (an eine Person gebunden). Teile dieses Wissens werden für andere Personen verfügbar gemacht durch Kodifizierung, das heißt beispielsweise durch Verschriftlichung des Inhaltes, ein anderer Teil ist ‚eingeschrieben‘ in die Fachpersonen (*tacit knowledge*), die am Erarbeitungsprozess neuen Wissens beteiligt sind.

Den letzten Schritt im DIKW-Prozess (Data-to-Information-to-Knowledge-to-Wisdom) bildet die Weisheit (Abbildung 3). Sie umfasst das „Verständnis von Grundsätzen“.

Abbildung 3: DIKW-Prozess (Clark, 2012)



Gefragt sind Informationen

Gefragt sind in der im vorliegenden Projekt geplanten Anwendung damit eigentlich nicht die Daten, sondern Informationen im Sinne von aufbereiteten, strukturierten Daten in nutzbarer Form und aus diesen generierten Erkenntnissen.

Big Data Analytics

Angesichts der in Kapitel 3.1.1 eingeführten Charakteristika von Big Data stellt gerade die Datenverarbeitung und die Gewinnung von Informationen eine sehr große, nicht nur technische Herausforderung für die Stadtentwicklungsplanung dar. Big Data-Analytics umfassen die Methoden zur möglichst automatisierten Erkennung und Nutzung von Mustern, Zusammenhängen und Bedeutungen (BITKOM 2014). Dabei existieren zahlreiche Methoden der Datenanalyse, welche sich, analog zu Big Data, schnell weiterentwickeln. Beispiele hierfür sind Data Mining, Visual Analytics, Maschinelles Lernen (Machine Learning) oder Text- und Bildanalytik (Text Mining, Computer Vision).

3.1.3 Typisierung

Gliederung in drei Typen

Wir gliedern Big Data in drei Typen (gemäß UNECE 2013), die sich hinsichtlich der Art der Datengenerierung unterscheiden:

Nutzergenerierte Daten: Die nutzergenerierten Daten entstehen durch Aktivitäten und oft insbesondere durch die Interaktion von Personen. Beispiele dieses Typs von Big Data sind Daten von klassischen sozialen Medien bzw. sozialen Netzwerken wie Facebook und Twitter, Daten von Fotosharing-Diensten wie Flickr und Daten von Wiki-Plattformen.

Transaktionsgenerierte Daten: Dieser Typ von Daten entsteht quasi als Nebenprodukt im Rahmen von ökonomischen Transaktionen. Klassische Beispiele sind Unternehmensdaten über Kunden und Umsätze, Daten von Tausch-, Immobilien- und Mobilitätsplattformen oder Daten aus dem Onlinehandel.

Sensorgenerierte Daten: Diese Daten entstehen nicht direkt durch Aktivitäten natürlicher oder juristischer Personen, sondern sie werden durch verschiedenartige Sensoren oder Sensornetzwerke erzeugt. Beispiele von sensorgenerierten Daten sind anhand von Induktionsschleifen und Seitenradaren erzeugte Verkehrsdaten, Daten aus Überwachungskamera-Systemen oder Mobilfunkdaten.

Keine trennscharfe Abgrenzung

Die drei Typen von Daten sind nicht trennscharf. Nicht alle Daten lassen sich zweifelsfrei und ausschließlich nur einer Kategorie eingliedern. Im Folgenden werden Mischformen vereinfachend einer Kategorie zugeordnet.

3.1.4 Abgrenzung: Big Data und Small Data

Small Data als Antithese

Die zum Teil unscharfe Definition von Big Data sowie der Hype, der um Big Data-Themen entstanden ist, führten dazu, dass „Big Data“ oft als Trendwort (miss)braucht wird. Vor diesem Hintergrund wurde der Begriff „Small Data“ geprägt. Ironischerweise legt auch dieser Begriff zumindest sprachlich den Fokus auf den Umfang von Daten. Der Begriff „Small Data“ ist insofern nur teilweise eine Antithese zu Big Data als dass er auch

für von Big Data abgeleitete Daten (oder eher: Informationen) und daraus resultierende Einsichten verwendet wird. Anderen dient der Begriff „Small Data“ aber als Erinnerung, dass es auch qualitativ sehr wertvolle Daten bzw. Datenquellen gibt, die nicht die typischen Merkmale (4 Vs) von Big Data aufweisen.

Abbildung 4: Big Data und Small Data (Kitchin 2015)

	Small data	Big data
Volume	Limited to large	Very large
Exhaustivity	Samples	Entire populations
Resolution and identification	Course & weak to tight & strong	Tight & strong
Relationality	Weak to strong	Strong
Velocity	Slow, freeze-framed/bundled	Fast, continuous
Variety	Limited to wide	Wide
Flexible and scalable	Low to middling	High

Abgrenzung

Neben Small Data gehören die folgenden Daten und Informationen nicht zu Big Data:

- (Amtliche) primärstatistische Erhebungen
- Sekundärstatistische Erhebungen (Nutzung bereits vorhandener Daten)
- (Nutzer-)Befragungen, Umfragen
- Statistiken zu (Förder-)Programmen
- Amtliche Geodaten
- Spezifische Foren (z. B. Beteiligungsprozesse im Web)
- Reine Informationsvermittlung (z. B. Info-Websites)

3.2 Crowd Data

Keine allgemeingültige Definition vorhanden

Der Begriff „Crowd Data“ verfügt über keine eindeutige Definition. Seine Bedeutung überlappt sich mit derjenigen verwandter Begriffe, die in der Literatur (zum Teil synonym) verwendet werden, zum Beispiel: User-Generated Content (UGC), Volunteered Geographic Information (VGI), Web 2.0, Mitmachweb, partizipatives Web, Produsage bzw. Produser. Auf diese Begriffe wird hier nicht näher eingegangen. Die Kenntnis dieser Begriffe ist aber nützlich für Recherchen zum Themenfeld.

Definition über die spezifische Art der Datenproduktion

Der Begriff „Crowd Data“ bezeichnet Daten, die aus Crowdsourcing stammen. Crowdsourcing ist das Auslagern einer Tätigkeit an eine oft sehr große Gruppe von Personen („Crowd“) (vgl. zum Beispiel Jeff Howe, 2006). Crowd Data sind folglich Daten, die von einer Gruppe von Personen oftmals dezentral und nach relativ informellen Regeln erarbeitet werden. Die wahrscheinlich bekanntesten Beispiele von Crowdsourcing sind Wikipedia, OpenStreetMap (OSM) und Amazon Mechanical Turk. Die Arbeit in einem Crowdsourcing-Projekt erfolgt in der Regel ohne Kompensation und freiwillig. Es sind aber auch monetäre oder nicht-

monetäre Anreize denkbar, beispielsweise der Status in der Community mittels Ranking der fleißigsten Teilnehmer*innen.

Citizen-Science-Daten

Unter dem Begriff „Citizen Science“ (Bürger*innen-Wissenschaft) wird eine Form der offenen Wissenschaft verstanden. Im Gegensatz zum Crowdsourcing werden Daten bei Citizen-Science-Projekten allerdings mithilfe von oder komplett von interessierten Laien gesammelt, beobachtet und/oder ausgewertet. Das Alleinstellungsmerkmal von Citizen-Science-Projekten gegenüber Crowdsourcing-Projekten im engeren Sinn besteht in stärkerer Formalisierung des Vorgehens, besserem Knowhow der involvierten Personen und einem engeren Bezug der Personen zur Thematik.

Crowd Data sind Big Data

Crowd Data zeichnen sich also durch eine spezifische Art der Datensammlung aus. Aufgrund ihrer Entstehungsmechanismen sind bei Crowd Data in der Regel alle vier für Big Data charakteristischen Vs (Volume, Variety, Velocity, Veracity) ausgeprägt.

4 Datenquellen

Kapitel 4 zeigt, wie die Quellen von Big Data kategorisiert wurden und beschreibt, welche Möglichkeiten die einzelnen Kategorien bieten, stadtraumbezogene Aussagen abzuleiten. Zudem werden die Datenquellen im Hinblick auf eine mögliche Nutzung hinsichtlich der Georeferenzierung der Daten sowie der Möglichkeit von Zeitreihen beurteilt.

Für stadtraumbezogene Aussagen ist in der Regel die Georeferenzierung der Daten entscheidend. Die Verknüpfung von georeferenzierten (Bewegungs-)Daten, die Ergänzung durch individuelle Attributdaten und kleinräumig relevante Standortdaten können wertvolle Rückschlüsse auf Verhaltensweisen und Abwägungsmuster der Nutzer*innen einer Stadt ergeben. Aus Sicht der Stadtentwicklungsplanung ist die Möglichkeit von Zeitreihen ein relevanter Faktor. Die Kurzlebigkeit und ständige Weiterentwicklung vieler Datenquellen machen jedoch Vergleiche über mittel- und langfristige Zeiträume fast unmöglich.

4.1 Kategorien von Datenquellen

Innerhalb der drei Typen von Daten – nutzergenerierte Daten, transaktionsgenerierte Daten und sensorgenerierte Daten – unterscheiden wir die folgenden Kategorien von Datenquellen.

4.1.1 Beschreibung der Datenquellen

Tabelle 1: Kategorien von nutzergenerierten Daten

Kategorie		*Beschreibung*
Fotoplattformen		Upload, Teilen und Kommentieren von Fotos
Ratingplattformen		Bewertung von Orten, Dienstleistungen und Aktivitäten (Restaurants, Unterkünfte, Kaffees, Attraktionen, Museen etc.)
Kartenportale		Abrufen von Informationen zum Raum (Points of Interest)
Social Media		Kommunikation online; Austausch von Nachrichten, Bildern, Filmen, Texten, Events etc.
GPS-Tracking		Aufzeichnen von Sportstrecken für Radfahrer*innen, Jogger*innen, Schwimmer*innen etc.
Suchportale		Suchen nach Informationen aller Art
Wikis		Sammeln und Austauschen von Wissen in Online-Enzyklopädien oder enzyklopädieähnlichen Nachschlagewerken

Datingportale		Suchen und Finden von Partner*innen / Beziehungen basierend auf verschiedenen Bedürfnissen
Businessnetzwerke		Knüpfen von geschäftlichen Kontakten und bilden von Netzwerken; Teilen von Beiträgen und Kommunikation

Tabelle 2: Kategorien von transaktionsgenerierten Daten

Kategorie		*Beschreibung*
Mobilitätsplattformen		Suchen und Finden von Mobilitäts- / Transportangeboten des ÖV und des privaten MIV
Immobilienportale		Suchen und Anbieten von Wohnungen, Gewerbeflächen, Gebäuden, Grundstücken zum Verkauf oder zur Miete / Pacht
Unterkünfte		Suchen und Buchen von Übernachtungen / Unterkünften
Ausschreibungs-plattformen		Suchen und Anbieten von Aufträgen zur Erbringung von Bau- und Dienstleistungen
Tauschplattformen		Tauschen und Teilen von Gegenständen und Dienstleistungen aller Art (exklusive Mobilität und Immobilien).
Jobbörsen		Suchen und Veröffentlichen von Stellenausschreibungen
Online-Handel		Suchen und Anbieten von neuen und gebrauchten Produkten durch Privatpersonen und Unternehmen
Unternehmensdaten		Unternehmensspezifisches Aufzeichnen von Daten zu Umsätzen, Kundenstruktur etc.
Mobilitätsdaten		Aufzeichnen von Fahrgastinformationen (Strecken, Umsätze, Kunden etc.)

Tabelle 3: Kategorien von sensorgenerierte Daten

Kategorie		*Beschreibung*
Navigationssysteme		Aufzeichnen von Fahrten mit Navigationssystemen
Mobilfunkdaten		Aufzeichnen von Personendaten, Orten von Mobilfunkaktivitäten von Telekomfirmen
Überwachungssysteme		Aufzeichnen von Filmmaterial zu verschiedenen Orten aus spezifischen Blickwinkeln

Gebäudedaten		Aufzeichnen von Leistungen / Aktivitäten aus Energieverbrauch, Ver- und Entsorgung, Emissionen
Biometrische Daten		Speicherung von Körpermerkmalen und Verhaltensweisen (Stimme, Gewicht, Fingerabdruck etc.)

Abgrenzungsfragen

Abgrenzungsfragen ergeben sich bei dieser Gliederung von Datentypen und Datenquellen-Kategorien insbesondere dort, wo Transaktionen über Plattformen abgewickelt werden (Unterkünfte, Wohnungen). Die resultierenden Daten erachten wir als eine Subkategorie der transaktionsgenerierten Daten, da die Daten im Rahmen einer Transaktion entstehen (z. B. Buchungs-, Reservierungsprozesse). Dies auch wenn manche dieser Plattformen Komponenten von sozialen Netzwerken aufweisen und die Transaktionen sich oftmals zwischen Nutzer*innen abspielen, nicht nur zwischen Unternehmen.

Keine abschließende Auflistung

Die Auflistung der Kategorien ist nicht abschließend. Es wurde versucht, die Auswahl und Beschreibung bereits mit einer „Stadtentwicklungsbrille“ vorzunehmen. Ein wesentlicher Fokus lag dabei in allen Typen von Daten auf den plattformgestützten Datenquellen. Dies geschah in der Annahme, dass die Daten einfacher zugänglich sind als sonstige unternehmensinternen Transaktionsdaten.

Vielzahl an unterschiedlichen Plattformen

Im Raster in Anhang A4 sind in den einzelnen Datenquellen-Kategorien jeweils ausgewählte Plattformen aufgeführt. Die Auflistung der Plattformen beschränkt sich auf die bekannteren. In den meisten Kategorien lässt sich eine Vielzahl an Plattformen finden.

4.1.2 Möglichkeiten stadtraumbezogener Aussagen

Grundsätzlich mögliche Aussagen aufzeigen

Im Folgenden ist für die einzelnen Kategorien von Datenquellen beschrieben, welche stadtraumbezogenen Aussagen sich grundsätzlich ableiten lassen. Diese Betrachtung erfolgt ohne Berücksichtigung der Zugänglichkeit der einzelnen Datenquellen. Da sich dies sehr dynamisch verändert, soll damit aufgezeigt werden, welche Aussagen grundsätzlich denkbar bzw. möglich sind.

Tabelle 4: Stadtraumbezogene Aussagen von nutzergenerierten Daten

Kategorie	*Stadtraumbezogene Aussagen*
Fotoplattformen	Ort der Aufnahmen gibt Hinweise auf Aufenthalte; Zeitpunkt der Aufnahmen gibt Hinweise zu zeitlichen Mustern (Tageszeiten, Wochentage etc.); Dichte von Aufnahmen gibt Hinweis auf Ausstrahlungskraft von Orten; Tagging der Fotos ermöglicht Suche nach Begriffen; Charakterisierung der Nutzer*innen aufgrund

	seiner Uploads (Touristen / Einheimische); Trajektorien zeigen Bewegung durch Stadt
Ratingplattformen	Ort der Bewertung gibt Hinweise auf Aufenthalte; Dichte von Bewertungen gibt Hinweis auf Ausstrahlungskraft; Kategorie der Bewertung gibt Hinweis auf Attraktivität; Zeitpunkt der Bewertung gibt Hinweise auf zeitliche Muster (Jahreszeiten), Charakterisierung der Nutzer*innen (Gliederung in Typen, z. B. nach Haushaltstyp oder Alter)
Kartenportale	Extraktion von Points of Interest (POI)
Social Media	Nennung bzw. Anzahl von Begriffen (z. B. Hashtags, Likes) gibt Hinweise auf aktuelle Entwicklungen / Trends / Aufmerksamkeit; Falls georeferenziert gibt der Ort Hinweise auf Aufenthalte; Häufigkeit und Art der Nutzung gibt Hinweise auf Nutzende
GPS Tracking	Strecken geben Hinweise auf Attraktivität von öffentlichem Raum / Straßen (für spezifische Nutzungen); Charakterisierung des Users aufgrund seiner Uploads (spezifische Profile); Erkennen von zeitlichen Mustern (Tageszeit, Woche, saisonal)
Suchportale	Nennung bzw. Anzahl von Begriffen gibt Hinweise auf aktuelle Entwicklungen / Trends / Aufmerksamkeit; Falls georeferenziert gibt der Ort Hinweise auf Aufenthalte bzw. bei Nutzerprofilen auch Charakterisierung der Nutzenden
Wikis	Extraktion von Points of Interest (POI)
Datingportale	Anzahl und Georeferenzierung von Nutzer*innen gibt räumliche Hinweise auf Haushaltsstrukturen und Lebensstile
Businessnetzwerke	Lebensläufe geben Hinweise auf Herkunft / räumliche Wanderungsbewegungen / Ausbildungsniveau / Lebensabschnitte; Vernetzungen geben Hinweise auf Wirtschafts- und Beziehungsnetzwerke; Geteilte Information / Begriffe geben Hinweise auf aktuelle Entwicklungen / Trends

Tabelle 5: Stadtraumbezogene Aussagen von transaktionsgenerierten Daten

Kategorie	*Stadtraumbezogene Aussagen*
Mobilitätsplattformen	Anzahl der Nutzer*innen der Angebote bzw. Anzahl Nutzungen gibt Hinweise auf Lebensstile; Standorte von Angeboten geben Hinweise auf räumliche Diffe-

	renzierung; zeitliche Muster (wann, wie lange); Charakterisierung der Anbieter; Charakterisierung der Nutzer*innen (Nutzerprofile: Häufigkeit, Orte, Zeiten)
Immobilienportale	Anzahl der Inserate (Verfügbarkeit) und Art der angebotenen Wohnungen / Gewerbeflächen (Kategorien) geben Hinweise auf Wohnraum- und Gewerbeflächenentwicklung; Spezifische Suchprofile (Kategorien) geben Hinweise auf Bedürfnisse / Nachfrage; Insertionsdauer gibt Hinweise auf Verhältnis Angebot / Nachfrage
Unterkünfte	Angebote: Orte, Anzahl und Preise; Suche: Bevorzugte Suchräume und Suchkriterien; Buchung: Anzahl der Buchungen gibt Hinweise auf Aufenthalte (Zeitpunkt, Dauer, Häufigkeit, Auslastung); Bewertung: gibt Hinweise auf Lagequalitäten (Lage, Erreichbarkeit, Lärm etc.); Charakterisierung der Nutzenden, in bestimmten Fällen Hinweise auf Zweckentfremdung von Wohnraum zu touristisch-kommerziellen Zwecken
Ausschreibungsplattformen	Anzahl der Ausschreibungen geben Hinweise auf (branchenspezifische) konjunkturelle Entwicklung; Nachfrage gibt Hinweise auf Konkurrenzsituation
Tauschplattformen	Ort, Zeitpunkt, und Art der angebotenen und genutzten Gegenstände gibt Hinweise auf aktuelle Nachfrage / Bedürfnisse. Gegebenenfalls Rückschlüsse auf die Nutzendenden (Lebensstil)
Jobbörsen	Anzahl der angebotenen Jobs und Nachfragen geben Hinweise auf konjunkturelle Entwicklung; Spezifische Suchprofile (Kategorien) geben Hinweise auf Nachfrage; Insertionsdauer gibt Hinweise auf Verhältnis Angebot / Nachfrage; Angebotene Jobs (Kategorien) geben Hinweise auf Entwicklung von Branchen und Anforderungsprofilen
Online-Handel	Art und Umfang der angebotenen / gekauften Artikel ermöglicht Rückschlüsse auf Nachfrage und Kaufkraft
Unternehmensdaten	Aussagemöglichkeiten abhängig von unternehmensspezifischen Daten
Mobilitätsdaten	Nutzungen geben Hinweise auf Bewegungsmuster

Tabelle 6: Stadtraumbezogene Aussagen von sensorgenerierten Daten

Kategorie	*Stadtraumbezogene Aussagen*
Navigationssysteme	Die aufgezeichneten Fahrten zeigen Belastung von Straßen / Orten und ermöglichen Rückschlüsse auf Stau sowie Luft- und Lärmbelastung; Über situative zeitliche Muster (Tageszeit, Woche, Saisonal) können zu Optimierungen im Verkehr und öffentlichen Räumen dienen; Entwicklungen / Auswirkungen von Infrastrukturplanungen können nachträglich gemessen werden
Mobilfunkdaten	Orts- und Bewegungsdaten zeigen Personendichte, Aufenthaltsorte, zeitliche Muster (Zeitpunkt, Zeitdauer). Pendlerströme, Fußgängerbewegungen, Verkehrsverhalten
Überwachungssysteme	Situative Informationen zu Personen- und Verkehrsdichte und -verhalten; Nutzung des öffentlichen Raumes
Gebäudedaten	„Verhalten" der Gebäude bietet Auskunft über räumliche Infrastruktur sowie Umwelt- und Klimasituation; Daten ermöglichen Rückschlüsse auf Verhaltensmuster von bestimmten Verbrauchergruppen; Grundlage für städtebauliche Energie-, Nachhaltigkeits- und Effizienzprojekte
Biometrische Daten	Erlauben über die Erfassung von „Gesundheitsparametern" im weitesten Sinne Aussagen zum Gesundheitszustand von Personen bzw. -gruppen

4.2 Einschätzung zu den Datenquellen

4.2.1 Georeferenzierung

Beispiele zur Nutzung von Big Data

Big Data wird bereits heute in verschiedener Weise zur Raumanalyse verwendet. Eine Sammlung von Beispielen, in denen nutzergenerierte Daten zur Raumanalyse eingesetzt wurden, findet sich in Anhang A2. In bisherigen Anwendungsbeispielen sind Fotoplattformen, Social Media und jegliche Kategorien, die mit GPS oder anderer Ortungstechnologie für Tracking arbeiten, wie z. B. Navigationssysteme oder Fitness-Apps, besonders verbreitet.

Georeferenzierung von entscheidender Bedeutung

Aus Sicht der Stadtentwicklungsplanung ist in der Regel die Georeferenzierung der Daten, bspw. der Nutzer*innen oder ihrer Aktivitäten, entscheidend. Je nach Plattform lassen sich unterschiedliche Informationen zu den Nutzer*innen herauslesen, etwa explizit bereitgestellte Attribute (Profildaten wie Alter, Wohnort, Geschlecht etc.) oder von ihrem Verhalten ableitbare (implizite) Eigenschaften (häufige Nutzer*innen, Aktivitä-

ten rund um spezifische Themen, Tages- oder Wochengang im Nutzungsverhalten etc.). Die Verknüpfung von georeferenzierten Bewegungsdaten, die Ergänzung durch individuelle Attributdaten und stadträumlich relevante Standortdaten können insbesondere als Längsschnittbetrachtung wertvolle Rückschlüsse auf Motivlagen und Abwägungsmuster der Nutzer*innen einer Stadt ergeben.

Stark generalisierte Aussagen

Die in Kapitel 4.1.2 dargelegten stadtraumbezogenen Aussagen in den einzelnen Kategorien sind generalisiert. Sie sind abhängig von den spezifischen Charakteristika (z. B. Regeln für die Teilnahme oder Aktivitäten auf der Plattform) bzw. der Zusammensetzung der Gesamtheit der Nutzer*innen einer Plattform. Die Unterschiede in der Aussagekraft und Nutzungsmöglichkeit zwischen unterschiedlichen Plattformen derselben Kategorie können dabei beträchtlich sein. Sie sind jedoch schwierig zu ermitteln. Es ist auch möglich, dass eine Plattform Eigenschaften aus mehreren Kategorien aufweist, zum Beispiel, wenn ein GPS-Tracking mit einem Rating verbunden wird.

Nutzer*innen bestimmen Ergebnis

Die Fotoplattform Geograph[3] (in Großbritannien, Irland und Deutschland) macht es sich zur Mission, zu jedem Kilometerquadrat im Land ein repräsentatives Foto zu sammeln (Motto: „*Photograph every grid square!*"). Nutzer*innen wetteifern, Fotos inklusive Standortangabe und Beschreibung hochzuladen. Wenn ein Bild die formalen Kriterien gut erfüllt, ist es ein sogenanntes „geograph" und die Urheberin oder der Urheber erhält Punkte. Diese Ausgangslage und das Anreizsystem machen Geograph zu einer Plattform, die nach ganz anderen Mechanismen funktioniert als die Fotoplattform Flickr. Abbildung 8 zeigt die resultierenden, stark unterschiedlichen Muster in den Daten: links die Dichte von geocodierten Fotos auf Flickr, rechts jene auf Geograph. Flickr fokussiert sich stark auf die innerstädtischen Räume, wo einerseits viele Personen leben und andererseits viel touristische Aktivität ist.

[3] http://www.geograph.org.uk

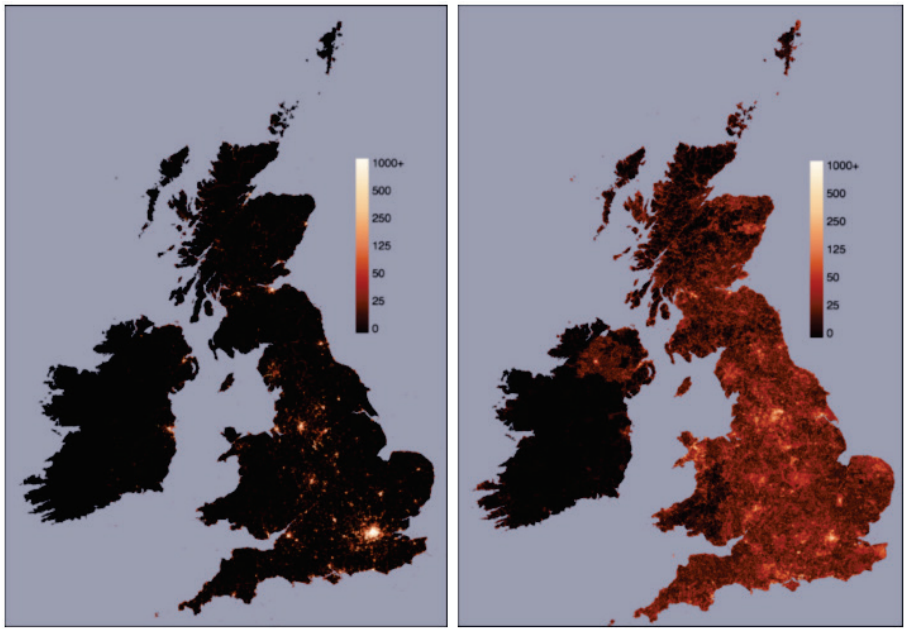


Abbildung 5: Dichte (Fotos pro m^2) von Flickr (links) und Geograph Fotos (rechts) (Purves et al., 2011)

4.2.2 Zeitreihen

Kurzlebigkeit

Die schnelle Entwicklung im Internetbereich und der rasche Entstehungsrhythmus neuer Angebote bergen die Gefahr, dass Datenquellen einer Kurzlebigkeit unterliegen können. Die Modalitäten der Datenproduktion und -abgabe können sich kurzfristig derart verändern, dass dies Auswirkungen auf Anwendungen von Big Data hat.

Plattformen im ständigen Wandel

Die Dynamik ist auf verschiedene Arten sehr hoch: So entstehen immer wieder neue Plattformen, während andere verschwinden oder zumindest an Popularität verlieren. Beispielsweise gab die bekannte Fotoplattform Panoramio[4] während der Bearbeitung dieses Berichts ihre Schließung per 4. November 2016 bekannt (und fiel damit aus dem Bewertungsraster).

Veränderungen bei der Erhebung von Daten

Neben den Plattformen selbst sind auch ihre Datenangebote einer dynamischen Entwicklung unterworfen. Beispielsweise ändern sich die Art der Datenerhebung oder die Möglichkeiten zum Datenbezug (vgl. Kapitel 5.3). Als ein Beispiel für veränderte Modalitäten bei der Datenerzeugung kann die Fotoplattform Flickr dienen. Flickr wurde in Forschung und Praxis wiederholt verwendet, um die räumliche Ausdehnung von unscharfen, nicht oder nicht nur offiziell definierten Toponymen (zum Beispiel „Norddeutschland“) und Räumen (zum Beispiel „Berliner Innenstadt“) zu bestimmen. Dazu wurden geocodierte Bilder mit ihren Metadaten von Flickr bezogen und deren Titel, Tags und Beschreibungen nach den gesuchten Begriffen gefiltert. Aus dem resultierenden Daten-

[4] http://www.panoramio.com

satz konnten dann Umrisse der gesuchten Ausdehnung abgeleitet werden (vgl. Abbildung 6).

Abbildung 6: Aus Flickr-Daten abgeleitete Regionen Londons (Hollenstein, Purves, 2010)

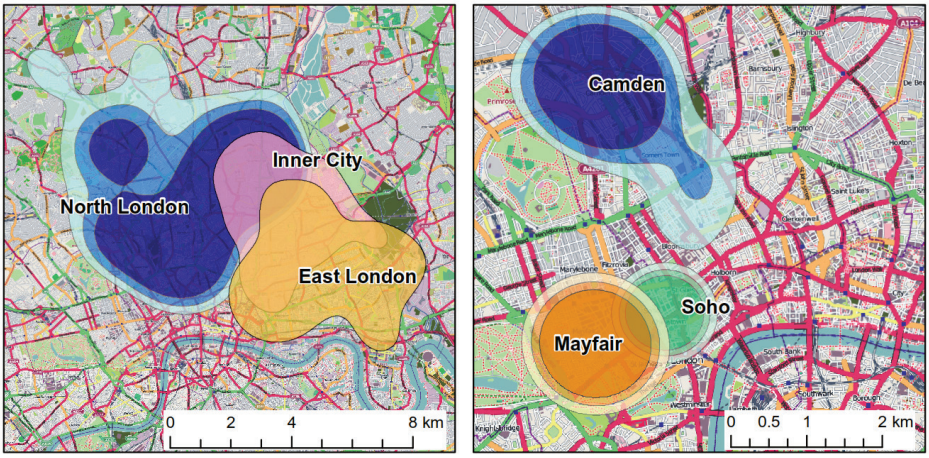


Figure 8: Generic ("northlondon," "innercity," and "eastlondon") and specific ("mayfair," "soho," and "camden") place names in London. For "innercity" and "eastlondon" only 50% contours of the volume surface are shown, for all other regions 50, 60, and 70%. London background mapping openstreetmap.org licensed under a Creative Commons Attribution-Share Alike License.

Änderung der Modalitäten schränkt Verwendungszeck ein

In der Zwischenzeit ist Flickr dazu übergegangen, Nutzer*innen beim Hochladen geocodierter Fotos selbst generierte geographische Tags anzubieten. Die Nutzer*innen entscheiden also nicht mehr (gänzlich) selbst, ob eine Fotografie bspw. in der „Berliner Innenstadt" aufgenommen worden ist. Dadurch würde die oben skizzierte Anwendung von Flickr-Daten aber (zumindest teilweise) zum Zirkelschluss; eine Änderung an den Modalitäten der Datenerzeugung hat somit die möglichen Verwendungszwecke der Daten eingeschränkt.

Dynamische Entwicklung der Funktionalitäten

Die einzelnen Plattformen entwickeln sich in ihren Funktionalitäten sehr dynamisch. Plattformen bieten ihre Angebote zunehmend über mehrere Kategorien an. So hat beispielsweise Airbnb sein Angebot um die Möglichkeit erweitert, Ausflüge oder Touren zu buchen. Mit der Integration einer Plattform, über die lokale Anbieter Touristinnen und Touristen ihre Dienste als Tourguides offerieren können, entwickelt sich das Unternehmen in Richtung eines umfassenden Reiseanbieters. Mit zusätzlichem Funktionsumfang oder veränderter Funktionsweise ändern sich aber auch die Nutzendenzahlen und -strukturen mit entsprechenden Verzerrungseffekten.

Dominanz großer Anbieter

Die Verbindung verschiedener Angebote wird insbesondere durch große Anbieter wie etwa Google vorangetrieben. Sie wollen ihre hohen Nutzerfrequenzen für unterschiedliche Dienstleistungen nutzen und auf diese Weise eine bedeutende Marktstellung erreichen.

Möglichkeit von Zeitreihen als kritischer Faktor für die Stadtentwicklungsplanung

Aus Sicht der Stadtentwicklungsplanung sind Zeitreihen von besonderem Interesse, weil damit Entwicklungsverläufe abgebildet werden könnten. Angesichts der Kurzlebigkeit und der ständigen Weiterentwicklung vieler Datenquellen, der damit implizit vorhandenen Abhängigkeit einer kontinuierlichen Datenverfügbarkeit zur Beschreibung bestimmter Phänomene, und aus bereits vorhandenen Erfahrungen ergibt sich, dass dieses Merkmal als ein besonders relevanter und kritischer Faktor einzuordnen ist. Abhängig von den jeweiligen Frage-/Themenstellungen ist einzuschätzen, ob daher eine Eignung für die Stadtentwicklungsplanung jenseits von „Moment-Aufnahmen" zu stadtentwicklungsplanerisch relevanten Themen gegeben ist.

5 Bezugskanäle

Kapitel 5 beschreibt die unterschiedlichen Bezugskanäle, also Wege, um Big Data von einem Anbieter für die Weiterverwendung in eigenen Untersuchungen zu beziehen. Dabei wird auf nutzergenerierte und plattformbasierte transaktionsgenerierte Daten fokussiert. Im Hinblick auf eine mögliche Nutzung erfolgt die Einschätzung zu den Bezugskanälen hinsichtlich der Zugänglichkeit der Daten, der Möglichkeit von Zeitreihen sowie der Datenqualität.

Die Zugänglichkeit der Daten wird durch die Tatsache erschwert, dass die meisten Daten mit einem bestimmten, meistens kommerziellen Zweck erhoben werden. Eine weitere Nutzbarmachung ist deshalb vielfach nur ein „Nebenzweck". Neben den Datenquellen sind auch die Bezugskanäle von Kurzlebigkeit geprägt. Dies beeinträchtigt Analysen über längere Zeiträume. Die Datenqualität ist für die Stadtentwicklungsplanung von besonderer Bedeutung, da die Daten als Planungs- und Entscheidungsgrundlagen verwendet werden. Die Qualität von Daten lässt sich aber ohne eine spezifische Prüfung kaum beurteilen. Die Qualitätsanforderungen sind zudem abhängig von der jeweiligen konkreten Anwendung.

5.1 Bezugskanäle im Überblick

Fünf Bezugskanäle für Daten

Im Folgenden werden die fünf hauptsächlichen Bezugskanäle für Big Data näher erläutert.

- **Programmierschnittstellen bzw. Application Programming Interfaces (APIs)**: Manche Datenproduzenten bieten APIs zur Programmierung von Apps für ihre Plattform an. Die Datenproduzenten werden so auch zu Anbietern ihrer eigenen Daten: Via publizierte APIs können Dritte auf Daten und Informationen zugreifen bzw. diese abfragen. Beispiele von Datenquellen mit APIs sind Twitter, Flickr und OpenStreetMap.
- **Webscraping**: Webscraping bezeichnet die Extraktion von Daten aus Webseiten. Webscraping macht sich die Struktur von Webseiten zu Nutze, um gezielt interessante Seiteninhalte zu finden und zu speichern. Dazu wird oft das Verhalten menschlicher Nutzer*innen simuliert (z. B. Klicken von Links, Ausfüllen und Absenden von Formularen, Scrollen).
- **Einkauf bei der Datenquelle**: Der Betreiber einer Datenquelle kann ein Interesse haben, die auf der eigenen Plattform erfassten Daten

abzugeben. Dies kann kommerziell motiviert sein oder auch der Imagepflege und der Etablierung als seriöse Organisation dienen[5].

- **Datenbroker**: Manche Datenquellen erlauben Datenbrokern, ihre Daten Dritten zugänglich zu machen. Das Angebot eines Datenbrokers kann auch neben der Verfügbarkeit einer API stehen (Beispiele von Datenbrokern sind Airdna (für Airbnb), Gnip (Twitter und diverse weitere Datenquellen für nutzergenerierte Daten) und Quintly.
- **Community**: Im Umfeld von Big Data und Crowd Data gibt es eine Community von Berufstätigen aber auch Enthusiasten und „Hobbyisten", die unregelmäßig veredelte Daten aus offenen Datenquellen bereitstellen. Zu der Community zählen beispielsweise Datenjournalist*innen, Forschende, Hackathon-Teilnehmende, das Startup-Ökosystem und die „Open"-Community. Die Community-Mitglieder bedienen sich bei ihren Aktivitäten typischerweise eines der anderen vier Bezugskanäle.

Beispiele im Anhang

In Anhang A6 sind ausgewählte Bezugskanäle (Plattformen und Datenbroker) hinsichtlich ihrer Zugänglichkeit, Qualität und räumlichen und zeitlichen Auflösung beurteilt.

5.2 Betrachtung der einzelnen Bezugskanäle

5.2.1 APIs

Vereinfachte Nutzung von APIs mit Wrappern

Oftmals existieren für die APIs bestimmter Datenproduzenten und ausgewählter Programmiersprachen sogenannte Wrapper. Dabei handelt es sich um Programmbibliotheken, deren Zweck die Vereinfachung der Formulierung von Abfragen an die API des Datenproduzenten ist. Beispielsweise ist *tweepy*[6] ein Wrapper für die API von Twitter in Python (einer Programmiersprache).

APIs in der Regel nicht als Vehikel zur Datenabgabe gedacht

Es muss berücksichtigt werden, dass die Verwendung von APIs aus Sicht von Datenproduzenten nicht primär den Interessen von Datenbeziehenden (also der Abgabe von Daten) dienen. Viele Datenproduzenten haben kein Interesse daran, sehr umfangreiches „Absaugen" ihrer Daten zu ermöglichen. Vielmehr stellen Datenproduzenten APIs bereit, um Dritten die Programmierung von Apps (z. B. eines eigenen Flickr-Clients) zu ermöglichen. Dies illustriert insbesondere die englische Bezeichnung „Ap-

[5] Beispielsweise wurde die Entscheidung von Uber, aggregierte hauseigene Daten Städten verfügbar zu machen, verschiedentlich als ein politisches Manöver interpretiert, das dazu dienen soll, die Legitimität von Uber zu stärken und Regulierung vorzubeugen; vgl. z. B. http://citiscope.org/story/2017/ uber-shares-mobility-data-urban-planners.

[6] https://github.com/tweepy/tweepy

plication Programming Interface" – also eine Schnittstelle für die Programmierung einer Applikation – sehr gut. Dieser Zielkonflikt zwischen Datenproduzenten und Interessenten an diesen Daten kann aus Sicht Letzterer zu Problemen bei der Verfügbarkeit führen.

Mögliche Beschränkungen von APIs

APIs können rechtlichen, technischen oder nutzergesteuerten Beschränkungen unterliegen. Diese Beschränkungen sind einerseits motiviert durch die oben beschriebene Interessenslage. Andererseits können Beschränkung auch (teil)motiviert sein durch den Schutz vor zu großen Kosten (Beschränkung der Abfragefrequenz erlaubt Serverkapazitäten enger auszulegen).

Limitierung von Abfragen

Bei den Einschränkungen ist das sogenannte Rate Limiting bzw. Query Quotas verbreitet: Dabei blockiert ein Anbieter Anfragen, wenn sie eine gewisse Frequenz (z. B. 2'000 Abfragen pro Tag, 100 Abfragen pro Viertelstunde) übersteigen. Damit verhindert ein API-Anbieter, das Dritte zu oft oder zu viele Abfragen tätigen und so die gesamte Informationsfülle (innerhalb kurzer Zeit) „absaugen" können und die technische Infrastruktur zu stark belasten. Eine in der Auswirkung ähnliche Beschränkung sind Response Quotas (beispielsweise bei Flickr eingesetzt) d. h. die Limitierung der Anzahl für eine bestimmte Anfrage zurückgelieferter Resultate.

5.2.2 Webscraping

Webscraping statt API

Webscraping (auch: Screen Scraping) hat sich in manchen Kontexten etabliert, um beispielsweise Beschränkungen einer API zu umgehen oder um die Inexistenz einer API zu überbrücken. Es wird beispielsweise angewendet, um die notwendigen Informationen für Vergleichsdienste (für Hotelpreise, Urlaubsreisen, u. ä.) zusammenzutragen.

Einschätzungen zur rechtlichen Zulässigkeit in Deutschland

Die Frage der rechtlichen Zulässigkeit von Webscraping tangiert in Deutschland, je nach konkreter Ausgestaltung des Scrapings, Bestimmungen des Urhebergesetzes (vgl. z. B. Solmecke 2013). Urheberrechtlich geschützte Daten dürften in der Regel nicht legal mit Scraping abgeschöpft und wieder veröffentlicht werden (Culbricht 2009). Allerdings muss für den Urheberrechtsschutz eine gewisse Schöpfungshöhe vorhanden sein. Erfüllen die Daten das Kriterium der Schöpfungshöhe nicht (z. B. reine Fakten wie Preise), kommt gemäß Culbricht (2009) das Kriterium zur Anwendung, dass „nicht wesentliche Teile" einer Datenbank bezogen werden dürfen.

Maßnahmen gegen Webscraping

Verschiedentlich wird auf die Möglichkeit von technischen Schranken und der Aufnahme entsprechender Klauseln in die Nutzungsbedingungen einer Webseite hingewiesen, um Webscraping zu verhindern oder zumindest zu erschweren (z. B. Solmecke 2013). Allerdings kommt laut Askan (2014) der Bundesgerichtshof in einem Urteil von 2014 zum

Schluss, dass das alleinige „Hinwegsetzen über den ausdrücklichen Willen [des Datenbankbetreibers], ein Screen-Scraping zu unterlassen" nicht genüge, um Unlauterkeit des Scrapings zu demonstrieren. Beim Überwinden technischer Schutzvorrichtung ist die Beurteilung aber nicht klar (ibid.). Solmecke (2013) hält dem entgegen, dass es 2013 in der Rechtsprechung „[w]eitgehender Konsens" ist, „dass – wenn weder wesentliche Teile der Datenbanken kopiert werden, noch es zur technischen Überlastung der ‚gescrapten' Seite kommt – automatisiertes Sammeln von Daten zulässig ist, solange die Seite rechtlich und technisch frei zugänglich ist."

Sorgfältige Prüfung

Es finden sich widersprüchliche Aussagen zu den Rahmenbedingungen der Legalität von Webscraping. Dessen Zulässigkeit ist folglich am besten für den Einzelfall abzuklären. Solange die rechtliche Lage nicht eindeutig geklärt ist, erscheint es für eine öffentliche Verwaltung nicht opportun, Webscraping kommerzieller Angebote einzusetzen. Im Zusammenhang mit Datenbrokern kann sich mittelbar dieselbe Situation ergeben, nämlich wenn der Datenbroker nicht mit dem Betreiber einer Datenquelle zusammenarbeitet.

5.2.3 Einkauf bei der Datenquelle

Veräußerung von Daten

Aus kommerziellen Beweggründen sind manche Betreiber von Datenquellen bereit, Extrakte der eigenen Daten für externe Nutzende abzugeben. Hier wird in Gesprächen und Verhandlungen zu prüfen sein, ob ein Zugang zu Daten (vielleicht aggregiert) möglich ist.

Motive für kostenlose Datenabgabe

Manche Betreiber sind vermutlich vorwiegend aus politischen Gründen auch bereit, Daten kostenlos abzugeben. Kostenlose Kooperationen mit Betreibern von Datenquellen können aus verschiedenen Gründen problematisch sein: Wie bei manchen anderen Bezugskanälen fehlt in der Regel die Garantie, dass das Angebot kostenlos oder überhaupt verfügbar bleibt. Wenn eine Behörde beispielsweise ein Monitoring auf derart bezogenen Daten aufbaut, begibt sie sich in eine Abhängigkeit, die die anfängliche Kostenfreiheit aufheben kann. Zudem kann die für die Verwaltung kostenlose Zusammenarbeit mit Betreibern von Datenquellen letzteren eine politische Akzeptanz verleihen.

5.2.4 Datenbroker

Verkauf aufbereiteter Daten

Das Geschäftsmodell von Datenbrokern basiert darauf, Daten von Plattformen zu sammeln und diese in aufbereiteter Form weiterzuverkaufen. Je nach Plattform können sich die Daten und die Objekte, worauf sich diese beziehen, unterscheiden. Bezogen auf Personen können je nach Plattform neben Basisinformationen wie Alter, Geschlecht oder Wohnort auch sozio-demographische Daten wie ethnische Zugehörigkeit, Bil-

dungsniveau oder Tätigkeiten erfasst werden. Vor allem in der Werbebranche ist die Nachfrage nach solchen Daten naturgemäß groß.

Spezialisierung von Datenbrokern

Datenbroker spezialisieren sich häufig auf soziale Netzwerke als Datenquellen. In diesem Zusammenhang haben sich insbesondere die Anbieter DataSift und Gnip mit ihren umfangreichen Werkzeugen auf dem Markt etabliert. Die Datenbroker ermöglichen den Anwender*innen die Aggregation, Filterung und Aufbereitung von nutzergenerierten Daten aus sozialen Netzwerken wie Facebook, Twitter, YouTube und zahlreichen anderen Datenquellen. Dabei liegen je nach Plattform historische sowie Echtzeit-Datenbestände vor.

Marktnische von Datenbrokern

Das Alleinstellungsmerkmal eines Datenbrokers – insbesondere bei gleichzeitigem Vorhandensein einer API – ist der einfachere und günstigere Zugang zu bestimmten Standarddaten oder Standardauswertungen (Reports) verglichen mit der programmiertechnischen Nutzung der API. Im Gegenzug sind komplexe, fragestellungsspezifische Abfragen eventuell ganz ausgeschlossen oder müssten eigens bestellt werden und sind dann auch teurer als Standardauswertungen.

Zusammenarbeit des Datenbrokers mit dem Betreiber einer Datenquelle

Die Verhältnisse zwischen Betreibern von Datenquellen und Datenbrokern sind nicht immer transparent. Manche Datenquellen erlauben Datenbrokern explizit, ihre Daten Dritten zugänglich zu machen. Es ist aber auch denkbar, dass Datenbroker gegen den Willen des Betreibers einer Datenquelle, versuchen ein Angebot aufzubauen. Die längerfristige Verfügbarkeit eines solchen Angebots ist naturgemäß unsicherer.

Potenzielle Konflikte

Am Beispiel von Airbnb kann man vor diesem Hintergrund einen Zielkonflikt zwischen den Datenproduzenten und Datenbrokern im Markt identifizieren: Der Datenbroker Airdna bietet verschiedene Daten und Informationen zu Airbnb an. Airdna ist unabhängig von Airbnb, aber insofern auf das Wohlwollen von Airbnb angewiesen, als letzteres den Zugang zu den Daten erschweren oder gar verunmöglichen könnte. Airdna richtet sein Angebot deshalb an Tourismusorganisationen, Forschung, Hotelmanagements, Maklerfirmen, Vermietungsfirmen und den Immobilienhandel. Airdna möchte nicht, dass Airdna-Daten dazu genutzt werden, Airbnb zu schädigen, zum Beispiel indem das lokale Airbnb-Angebot im Hinblick auf ein Verbot analysiert wird.[7]

5.2.5 Community

Zusammensetzung der Community

Die Community rund um Big Data umfasst beispielsweise Datenjournalist*innen, Forschende, Hackathon-Teilnehmende, das Startup-

[7] Quelle: Aussage aus einem Telefongespräch zwischen Airdna und EBP im Oktober 2015

Ökosystem und die Open-Gemeinschaft. Mitglieder der Community nutzen für den Bezug von Big Data typischerweise einen der anderen Bezugskanäle, oftmals APIs.

Lokalwissen von Vorteil

Die Zusammensetzung der Community fluktuiert, was auch ihre Aktivitäten und deren Langlebigkeit negativ beeinflussen kann. Besonders interessant kann die Zusammenarbeit mit einer lokalen Community sein, da diese mit den lokalen Gegebenheiten besser vertraut ist und deshalb Daten passender für den lokalen Kontext aufbereiten und weiterverarbeiten kann als ein potenziell global tätiger Betreiber einer Datenquelle oder große Datenbroker.

Geben und Nehmen

Für eine erfolgreiche Zusammenarbeit mit einer Community sind Elemente des Community Managements zu beachten. Wichtig ist zum Beispiel die Kommunikation auf Augenhöhe und eine wechselseitige Beziehung, bei der auch die Community in irgendeiner Form von der Zusammenarbeit profitiert.

5.3 Einschätzung zu den Bezugskanälen

5.3.1 Zugänglichkeit

Zugänglichkeit nicht pauschal zu beurteilen

Eine pauschale Beurteilung der Zugänglichkeit der Daten ist schwierig vorzunehmen. So muss etwa bei einer API geprüft werden, inwieweit die Limitierungen einen konkreten Anwendungsfall be- oder verhindern könnten. Diese Beurteilung kann im notwendigen Detailgrad in der vorliegenden Studie nicht geleistet werden; sie sollte bei einer Anwendung einer API, beispielsweise bei der Umsetzung einer prototypischen Anwendung, plattformspezifisch geprüft werden.

Daten als kommerzielles Gut

Es kann jedoch festgehalten werden, dass die Zugänglichkeit mit verschiedenen Herausforderungen verbunden ist. Die Daten werden mit einem bestimmten, meistens kommerziellen Zweck erhoben. Eine weitere Nutzbarmachtung ist deshalb oftmals nur ein „Nebenzweck". Für viele Betreiber von Datenquellen sind ihre Daten ein sehr wichtiges oder gar das zentrale Gut, um das sich ihr Geschäftsmodell aufbaut. Je wichtiger die Daten für den Erfolg des eigenen Geschäftsmodells sind, desto geringer wird in der Regel die Zugänglichkeit sein.

My Data: Eigene Daten freigeben

Entgegen der Herangehensweise von Datenbrokern, bei denen die Nutzer*innen häufig ahnungslos darüber sind, dass Informationen über sie gesammelt werden, gibt es zum Teil Anbieter, bei denen die Daten nur mit der Zustimmung der Nutzer*innen generiert bzw. gesammelt werden. Die Idee dahinter ist, dass Nutzer*innen selbst über die (Weiter-) Verwendung ihrer Daten bestimmen. User*innen können dabei angeben, welche Informationen sie preisgeben möchten. Kunden werden dann

diese Informationen in anonymisierter Form zur Analyse angeboten und die Nutzer*innen werden an den Einnahmen beteiligt. Der Vorteil darin liegt zum einen in der Verlässlichkeit und Aktualität der Daten und zum anderen in der Transparenz den Nutzer*innen gegenüber. Dieses Konzept firmiert unter dem Begriff „My Data“[8] bzw. in Großbritannien auch „midata“[9]. Trotz der zahlreichen attraktiven Ideen nehmen derartige Lösungen auf dem Markt für Daten bisher eine marginale Rolle ein.

Politischer Druck verschlechtert die Datenverfügbarkeit

Neben den kommerziellen Interessen von Betreibern möglicher Datenquellen ist ein weiterer Grund für eine manchmal schlechte Verfügbarkeit von Daten der regulatorische Druck: Manche Datenproduzenten haben ein Geschäftsmodell, das von offiziellen Stellen kritisch beurteilt wird. Prominente Beispiele sind die Firmen Uber und Airbnb und ihre Angebote im europäischen Markt.

Datenschutzvorbehalte

Auf manchen Plattformen können die Nutzenden selbst bestimmen, ob ihre Aktivitäten auf der Plattform privat sein sollen oder nicht. Private Twitterkonten (sogenannte „Protected Accounts“) sind dann beispielsweise via API nicht abfragbar. In der Regel werden Informationen zu solchen Nutzer*innen nicht in den irgendwie erhältlichen Daten auftauchen, egal über welchen Kanal diese bezogen werden.

5.3.2 Zeitreihen

Kurzlebigkeit kann Längsschnitt-Untersuchungen erschweren oder verhindern

Die hier betrachteten Bezugskanäle sind zum Teil ähnlich wie die Datenquellen eher kurzlebig (vgl. Kapitel 4.2.2). Bezogen auf Bezugskanäle bedeutet Kurzlebigkeit zum Beispiel, dass sich die Modalitäten der Nutzung einer API, der Umfang angebotener Daten oder der Preis eines Angebots verändern können. Die Kurzlebigkeit von Bezugskanälen, aber auch von Datenquellen selbst, vermindert die durchschnittliche Längsschnittfähigkeit von Big Data-Angeboten.

5.3.3 Datenqualität

Eingeschränkte Datenqualität

Die Qualität der über die diversen Bezugskanäle erhältlichen Daten gilt es kritisch zu beurteilen, wie dies im Raster „Bezugskanäle“ für einige Vertreter gemacht worden ist. Die relevanten Qualitätskriterien dazu sind (vgl. auch Anhang A6):

[8] Vgl. https://opendata.ch/swiss-data-alliance und http://opendataconsulting.ch/wp-content/uploads/_2015/08/Ambassador-Club_Digitale_Schweiz_20160712.pdf für My Data-Aktivitäten im Rahmen der Schweizerischen Datenpolitik auf Bundesebene.

[9] Vgl. zum Beispiel https://www.gov.uk/government/news/the-midata-vision-of-consumer-empowerment und http://www.gocompare.com/money/midata

- Objektivität: Sind die zu generierenden Informationen repräsentativ (genug)? Bestehen Verzerrungen bspw. in der Gruppe der Nutzer*innen? Hier muss angemerkt werden, dass das Interesse an „Verzerrungen“ manchmal genau Anlass sein können bspw. nutzergenerierte Daten zu untersuchen (bspw. Untersuchung von Instagram oder Foursquare, um die Aktivitäten vergleichsweise junger Personen charakterisieren zu können).
- Längsschnittfähigkeit (Nachhaltigkeit)
- Qualitätssicherung: Nennt der Datenlieferant bspw. explizit unternommene Schritte zur Qualitätssicherung?

Im Allgemeinen lässt sich die Qualität unbesehen, d. h. vor Bezug von Daten oder – idealerweise dem Bezug und der Analyse von Testdaten – bei den meisten Betreibern von Datenquellen nicht sehr gut beurteilen.

Anwendungsspezifische Beurteilung der Qualität

Die Qualität von Daten sollte möglichst vor dem Hintergrund einer konkreten Anwendung bzw. einem geplanten Einsatz geprüft werden (Stichwort: „fitness for use“). Ein bestimmter Qualitätsaspekt kann im einen Anwendungsfall unkritisch sein, einen anderen Anwendungsfall aber verunmöglichen.

6 Nutzungspotenzial für die Stadtentwicklungsplanung

In Kapitel 6 werden die generellen Nutzungspotenziale der drei Typen von Big Data sowie von Crowd Data für die Stadtentwicklungsplanung betrachtet und mit Beispielen illustriert.

Nutzergenerierte Daten vermitteln den Eindruck, dass damit „reale" Verhaltensweisen und Präferenzen ablesbar sind. Die Informationen erfordern jedoch eine Interpretation von Motivlagen, was die Gefahr von Fehl- oder Überinterpretation mit sich bringt. Verschiedene Hürden erschweren zudem heute noch eine verstärkte Nutzung der Daten. Die Nutzung von *transaktionsgenerierten Daten* ist abhängig vom Geschäftsmodell der jeweiligen Anbieter. *Sensorgenerierte Daten* besitzen angesichts großer und immer weiter steigender Datenmengen sowie aufgrund ihrer Validität ein beträchtliches Potenzial. Ihre Analyse und Nutzung erfordert jedoch umfassende Kompetenzen und Ressourcen. *Crowd Data* bzw. *Crowdsourcing* bieten eine neue bzw. ergänzende Möglichkeit einer gezielten und unmittelbaren Interaktion mit der Stadtbevölkerung.

6.1 Nutzungspotenzial von nutzergenerierten Daten

Nutzergenerierte Daten vermitteln „menschliche Nähe"

Nutzergenerierte Daten scheinen für die Stadtentwicklungsplanung im Grundsatz interessant. Da sie von Bewohner*innen generiert werden, vermitteln sie den Eindruck von „menschlicher Nähe" und suggerieren, dass damit „reale" Verhaltensweisen und Präferenzen ablesbar sind. Die Anwendung von durch neue Technologien nutzergenerierten (Massen-) Daten in der Stadtentwicklungsplanung ist jedoch wenig verbreitet und hat etwas stark Pionierhaftes, mit entsprechender Öffentlichkeitswirkung.

Zugänglichkeit und Aussagekraft ist limitiert

Es bestehen verschiedene Hürden, um die Daten effektiv nutzen zu können:

- Die Zugänglichkeit zu nutzergenerierten Daten ist eingeschränkt. Vielfach lassen sich nur generelle Aussagen treffen. Detailliertere Informationen wären zwar vorhanden, sind aber auch aus Gründen des Datenschutzes oder den geschäftlichen Interessen der Datenproduzenten nicht verfügbar.
- Nutzergenerierte Daten sind häufig sehr unstrukturiert (Fotos, Text etc.). Sie benötigen entsprechend eine aufwändige datentechnische und methodisch-analytische Aufbereitung.

- Die Qualität der Daten ist vielfach beschränkt bzw. lässt sich kaum beurteilen. Der Umgang mit Verzerrungen gegenüber der räumlich zurechenbaren Grundgesamtheit ist noch weitgehend unklar. Auch die Einflüsse durch falsche Identitäten und soziale Roboter (Social Bots, Sock Puppets) sind noch weitgehend ungeklärt. So gehen Schätzungen davon aus, dass bis zu 20 Prozent aller Twitter-Accounts schon heute Social Bots sind.
- Zeitreihen sind sehr schwierig zu erstellen. Die hohe Volatilität der Plattformen mit schnell wechselnden Funktionalitäten macht Vergleiche über mittel- und langfristige Zeiträume fast unmöglich.

Interpretation von Motivlagen notwendig

In inhaltlicher Hinsicht besteht die zentrale Herausforderung darin, die Informationen zu interpretieren und auf Verhaltensweisen zurückzuschließen. Nutzergenerierte Daten lassen unter Umständen eine Mustererkennung via Korrelationen im Raum zu. Weshalb diese Nutzertypen etwas und wo tun, entzieht sich diesem Zugang (keine Motivforschung, Revealed Preference). Diesbezügliche Interpretationen sind aufgrund von Fehl- oder Überinterpretation entsprechend mit Unsicherheiten behaftet (Ruths, Pfeffer 2014). Zusätzliche Hinweise zur Interpretation lassen sich gewinnen, wenn neben der eigentlichen Nutzung (z. B. Hochladen eines Fotos, einloggen in einer Örtlichkeit) weitere Informationen damit verbunden sind, wie etwa Begriffe (Tags), Bewertungen oder Kommentare. Die Auswertung solcher Hinweise ist jedoch mit einem hohen Aufwand verbunden.

Überlagerung individueller Muster

Im Bereich Fotoplattformen und Social Media sind insbesondere Flickr und Twitter bereits häufig für Analysen verwendet worden. Das liegt wahrscheinlich an der verhältnismäßig guten Zugänglichkeit der Daten dieser Quellen via APIs (application programming interface). Aus der Überlagerung von individuellen Nutzungs- bzw. Verhaltensmuster im Raum lassen sich Aussagen zur Anzahl bzw. zur Dichte sowie der Mischung der Nutzer*innen an spezifischen Orten gewinnen.

Beispiel: Flickr-Daten

Abbildung 7 und Abbildung 8 zeigen eine Beispiel-Analyse von Flickr-Daten (Straumann et al. 2014). Dabei wurden geocodierte Fotos und deren Metadaten via API von Flickr bezogen. Die Fotostandorte sowie die zugeordneten Flickr-Benutzer*innen wurden verschiedentlich analysiert. Basierend auf dem dem Flickr-Konto hinterlegten Herkunftsland wurden Tourist*innen von Einheimischen unterschieden. Abbildung 7 zeigt, wie sich die von den beiden Gruppen (stark) fotografierten Gebiete unterscheiden. Abbildung 8 zeigt eine Analyse von Trajektorien, also die Wege von Personen, die sich aus verschiedenen, zeitlich relativ nah aufeinanderfolgenden Fotostandorten rekonstruieren lassen. Dabei zeigt sich beispielsweise, dass die beiden Personengruppen unterschiedliche Gebiete auf leicht unterschiedlichen Pfaden besuchen bzw. dass manche

Pfade von der Gruppe der Tourist*innen gar nicht oder deutlich weniger begangen werden.

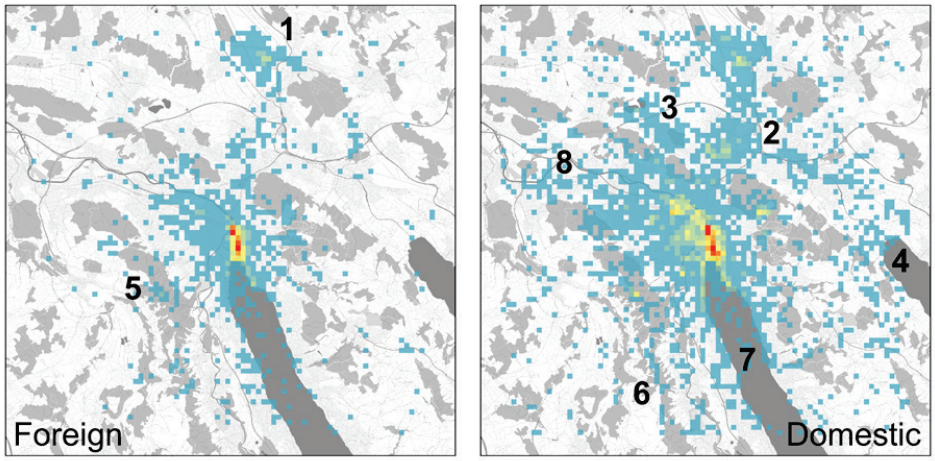


Abbildung 7: Dichte von aus- und inländischen Fotografen in der Gegend von Zürich, basierend auf Flickr-Daten (Straumann et al. 2014).

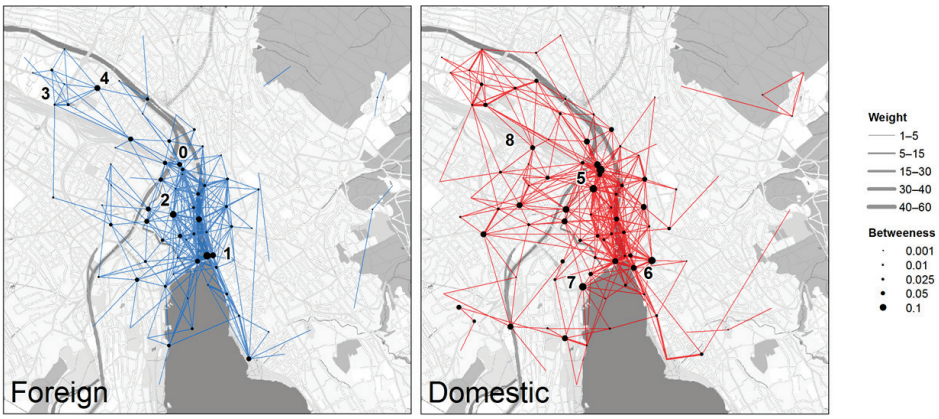
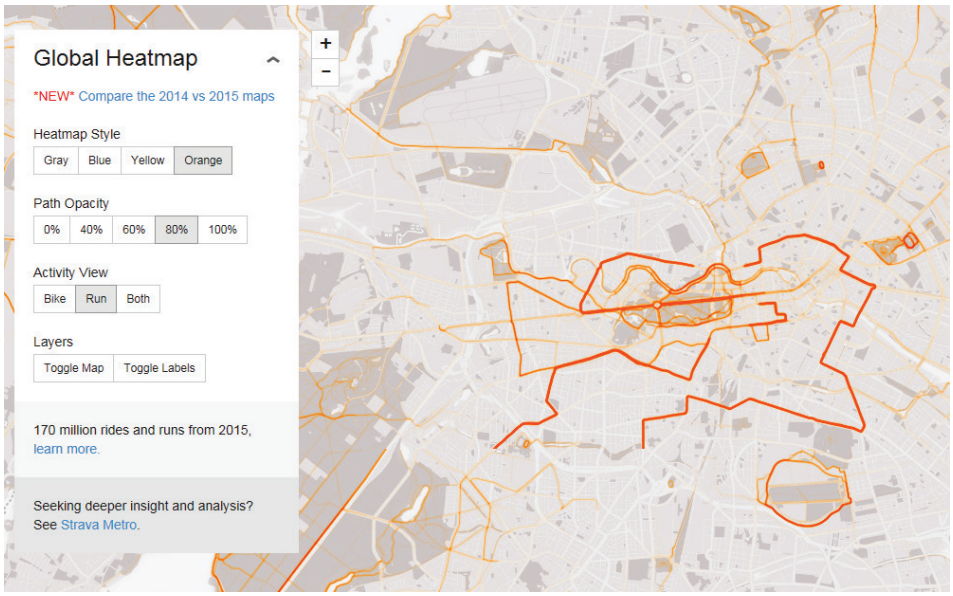


Abbildung 8: Netzwerk aus Trajektorien von aus- und inländischen Fotografen in der Stadt Zürich, basierend auf Flickr-Daten (Straumann et al. 2014).

Beispiel: Strava Heatmap

Strava ist ein soziales Netzwerk für Sportler. Die Nutzer*innen dokumentieren dabei ihre sportlichen Aktivitäten, insbesondere Rad- und Laufaktivitäten. Die gesammelten und visualisierten GPS-Tracks werden von Strava als sogenannte Heatmaps kostenlos zur Verfügung gestellt (Abbildung 9). Mit dem kostenpflichtigen Angebot „Strava Metro“ erhält man Zugang zu den dahinterliegenden Daten. Darauf basierend sind detaillierte Analysen zum Verhalten auf Arbeits- oder Freizeitwegen möglich.

Abbildung 9: Strava Global Heatmap[10]



Ein grobes Bild vermitteln

Trotz der Einschränkungen können nutzergenerierte Daten einen Eindruck eines Sachverhaltes vermitteln und Hinweise auf näher zu betrachtende Zusammenhänge oder Fragestellungen liefern. Dies insbesondere in Kombination bzw. Überlagerung mit weiteren Datenquellen. Damit lassen sich etwa Aussagen zu allgemein gefassten Themen von Stadtentwicklungsplänen wie der Umfeld- oder Standortqualität herstellen. Diese sind jedoch in den meisten Fällen nur wenig konkret.

Kurzfristige Veränderungen erkennbar machen

Aus nutzergenerierten Daten lassen sich auch Veränderungen über die Zeit ablesen. Angesichts der Schnelllebigkeit der Plattformen und ihrer dynamischen Entwicklung umfasst dies jedoch nur kurzfristige Trends. Langfristige Entwicklungsrichtungen dürften kaum zuverlässig aufzuzeigen sein.

Große Zahl von Nutzern notwendig

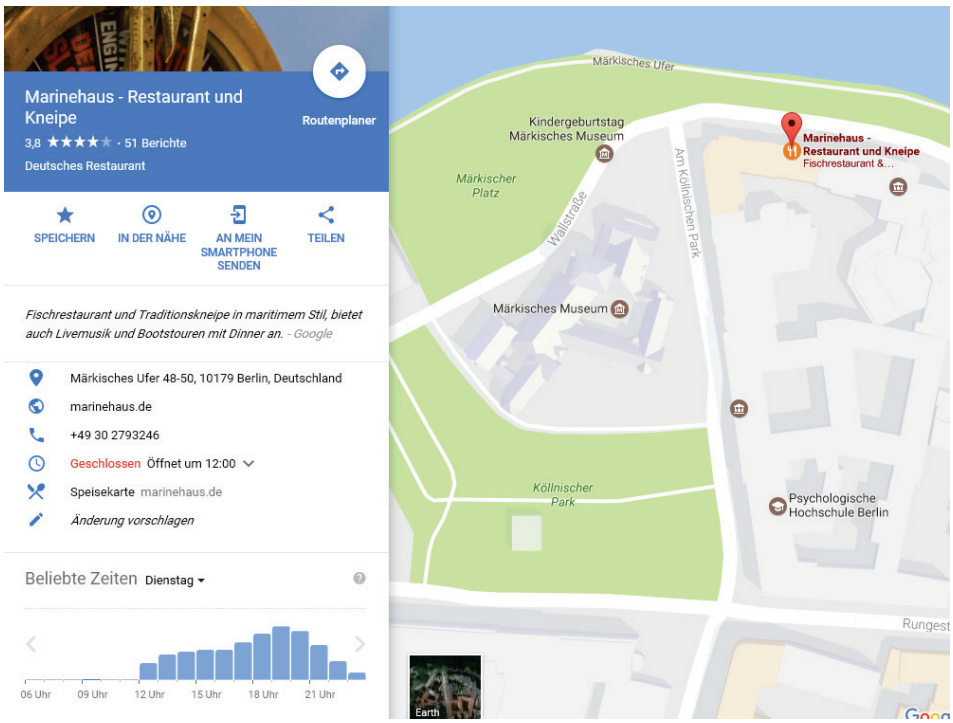
Um die Probleme mit der Datenqualität zu reduzieren, braucht es eine ausreichend große Anzahl von Nutzer*innen. Entsprechend sind nur Plattformen von Interesse, die mit großen Nutzerzahlen über einen längeren Zeitraum präsent sind. Gleichzeitig stellt sich auch hier wiederum die Frage der Repräsentativität der Nutzenden in Relation zur planerisch-konzeptionellen Zielgruppe bzw. der Allgemeinheit als Adressat von stadtentwicklungsplanerischen Konzepten.

Beispiel: Google Places

[10] http://labs.strava.com/heatmap/#12/13.39972/52.50149/orange/run [Zugriff: 10.12.2016]

Google Maps ist eine der weltweit meistgenutzten Apps. Seit 2016 zeigt die Karten-App auch die Stoßzeiten von Geschäften und anderen Lokalitäten. Sie stützt sich dabei auf die Auswertung der Standortdaten von Nutzer*innen, welche diese zur Nutzung freigegeben haben.

Abbildung 10: Google Places[11]



Herausgreifen spezifischer Einzelinformationen

Aus den nutzergenerierter Daten können auch spezifische Einzelinformationen herausgegriffen werden. Beispielsweise können einzelne Fotos, Kommentare oder Bewertungen zu ganz spezifischen Räumen (z. B. zur Nutzung eines Parks) Hinweise auf ganz konkrete Themen oder Fragen liefern, welche sich vertieft bearbeiten lassen.

6.2 Nutzungspotenzial transaktionsgenerierter Daten

Starke Unterschiede zwischen den Kategorien

Die Kategorien innerhalb des Typus der transaktionsgenerierten Daten weisen ganz unterschiedliche Charakteristika auf und werden deshalb im Folgenden hinsichtlich ihrer Nutzungspotenziale spezifisch betrachtet Sie lassen sich gliedern in

- Plattformbasierte Transaktionen

[11] http://google.maps.ch [Zugriff: 14.02.2017]

- Transaktionen privater Unternehmen
- Transaktionen öffentlicher Unternehmen

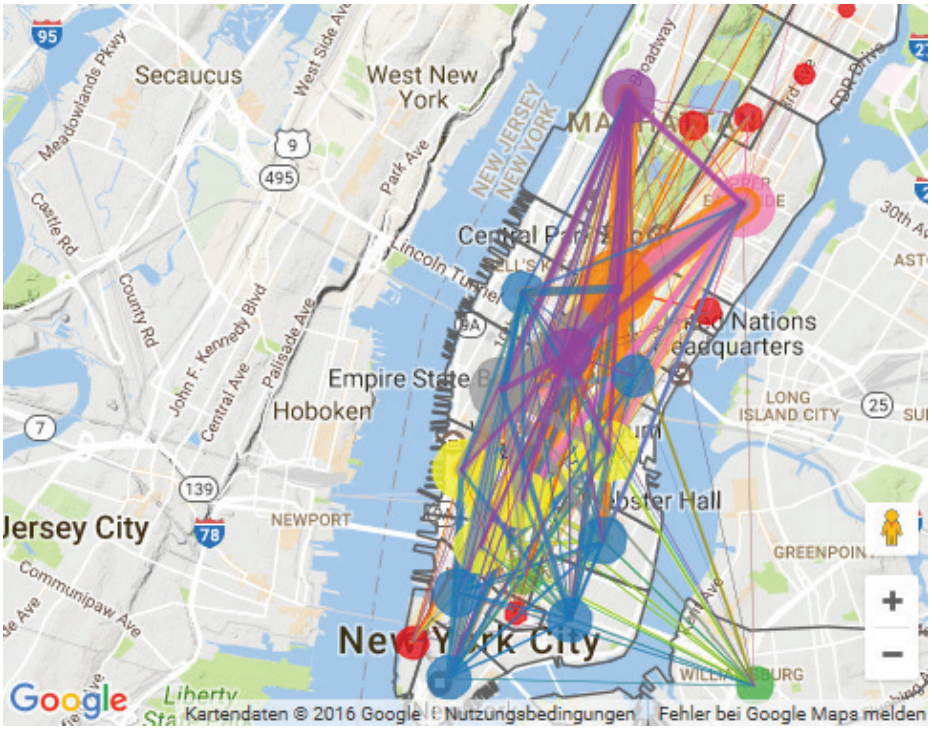
Plattformbasierte Daten

Die plattformbasierten Daten weisen in Teilen vergleichbare Charakteristika auf wie die nutzergenerierten Daten. Auch bei privaten Unterkunftsangeboten oder Immobilienportalen werden die Daten durch eine Vielzahl von Nutzer*innen generiert. Da es sich jedoch um geschäftsbezogene Informationen handelt, kann davon ausgegangen werden, dass die Zuverlässigkeit, die Genauigkeit und damit auch die Aussagekraft der Informationen höher einzuschätzen ist als bei den nutzergenerierten Daten.

Beispiel: Uber

Ein Beispiel für transaktionsgenerierte Daten ist der Fahrdienstanbieter Uber. Diese Daten können auch visualisiert werden. So lässt sich beispielsweise darstellen, wie stark einzelne Quartiere über Uber-Fahrten miteinander „verbunden" sind.

Abbildung 11: Analyse von Uber-Daten[12]



Beispiel: Uber Movement

Das Angebot „Uber Movement" (lanciert am 8. Januar 2017) ist ein Beispiel einer Kooperation der Mobilitätsplattform Uber mit Planer*innen von Städten. Es zeigt einige der oben angesprochenen problematischen Punkte: Die Datenabgabe erfolgt zumindest vorerst kostenlos. In der Ankündigung des neuen Angebots ist allerdings keine Aussage zur künftigen Kostenfreiheit enthalten. Das Angebot ist zudem als Public Relati-

[12] newsroom.uber.com/uberdata-mapping-a-citys-flow-using-ubers-ridership-data/ [Zugriff: 10.12.2016]

on-Aktion von Uber zu werten, das in vielen Ländern von Regulatoren und Politiker*innen kritisch beurteilt wird und z. B. wegen Sozialversicherungsfragen in diverse Rechtsfälle verwickelt ist. Trotz der augenscheinlichen Kooperation mit Planer*innen weigert Uber sich aber beispielsweise, von New York City angefragte Daten abzugeben (Hatch, 2017). Die über Uber Movement verfügbaren Daten sind laut der National Association of City Transportation Officials der USA (NACTO) zudem nicht gut auf die Bedürfnisse von Städten abgestimmt und folgten nicht der „Best Practice" (ibid.).

Beispiel: Auswertung von Airbnb-Angeboten

Airbnb als bekannteste digitale Vermittlungsplattform für Privatunterkünfte spielt im Beherbergungsmarkt eine bedeutende Rolle. Sie beeinflusst jedoch auch den Wohnungsmarkt. Diese Debatte wird gerade in Berlin lebhaft geführt. In Berlin ist es seit 2014 gesetzlich verboten (Zweckentfremdungsverbot-Gesetz), eine Mietwohnung ohne Genehmigung als Ferienwohnung zu vermieten. Mittels Statistik-Software sind automatisierte Abfragen zu den Angeboten (nach Preis, Kategorie, Kapazität etc.) möglich (vgl. Kanton Zürich 2016). Die Website http://www.airbnbvsberlin.de macht sich diese Auswertungsmöglichkeiten zunutze und stellt den Berliner Airbnb-Markt dar.

Beispiel: Auswertung von Stellenanzeigen

Ein weiteres Beispiel zeigt die Analyse von Stellenanzeigen. Basierend auf der Datenbasis zu den Stellenangeboten des Unternehmens x28, welches offene Stellen in der Schweiz direkt auf den Websites der Firmen crawlt, wurde beispielweise eine Analyse zum Thema „Das Stellenangebot ein Jahr nach der Aufhebung des Mindestkurses des Schweizer Frankens" erarbeitet (Koellreuter 2016).

Private Unternehmensdaten

Die Verfügbarkeit privater Unternehmensdaten dürfte kaum gegeben sein. Sie wurden deshalb im Rahmen dieses Projektes auch nicht näher betrachtet. Für die Stadtentwicklungsplanung ist jedoch, im Dialog mit den Unternehmen, eine Nutzung in konkreten Planungen denkbar.

Öffentliche Unternehmensdaten

Nutzungspotenziale bestehen auch hinsichtlich öffentlicher Unternehmensdaten, beispielsweise bei den Mobilitätsanbietern (z. B. Berliner Verkehrsbetriebe). Die konkreten Nutzungsmöglichkeiten wären spezifisch abzuklären.

6.3 Nutzungspotenzial sensorgenerierter Daten

Große Datenmengen

Über Sensoren wird eine enorme Menge an Daten gesammelt. Sensorgerierte Daten sind auch dadurch charakterisiert, dass sie in (fast) Echtzeit zur Verfügung stehen. Mit dem Internet of Things (IoT) dürfte sich diese Dynamik weiter verstärken. Das „Internet der Dinge" bezeichnet dabei Datenfluss zwischen physischen Gegenständen (Dingen) übers Internet. Diese Gegenstände schicken Daten in ein Netzwerk und liefern

Informationen über einen Zustand des Gegenstands, welche in weiterer Folge verwertet werden können. Dabei wird auch diskutiert, wie das IoT für die Stadtentwicklungsplanung genutzt werden kann (z. B. in Wien, (https://www.digitaleagenda.wien/de).

Mobilfunkdaten

Relevante Informationen zur Analyse städtischer Strukturen können aus Mobilfunkdaten gewonnen werden. In den vergangenen drei Jahren sammelten und beforschte ein internationales Konsortium beispielsweise die anonymisierten Mobilfunkbewegungsdaten der Wiener Pendler (z. B. http://www.urbanapi.eu/). Auch für spanische Städte (Louai et al. 2014) oder chinesische Städte (Yuan, Raubal 2012) wurden räumliche Strukturen anhand von Mobilfunkdaten analysiert (Abbildung 13).

Abbildung 12: Räumliche Darstellung von Mobilfunkdaten (Yuan, Raubal, 2012)

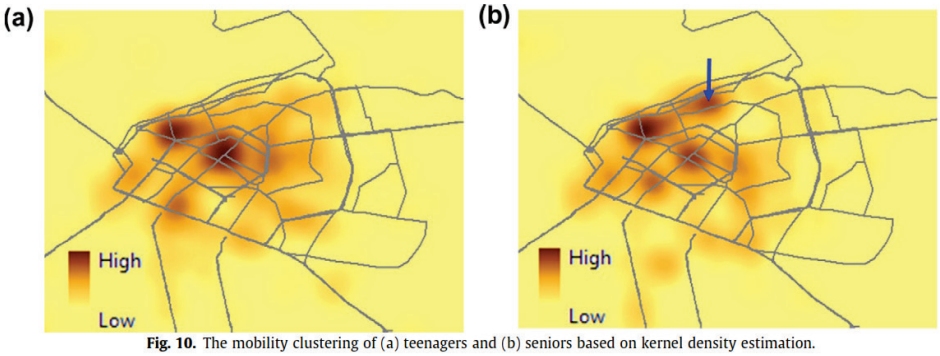


Fig. 10. The mobility clustering of (a) teenagers and (b) seniors based on kernel density estimation.

Einsatzmöglichkeiten in der Stadtplanung

Die sensorgenerierten Daten sind auch derjenige Typus, welcher am unmittelbarsten mit der Stadtplanung verbunden ist. Dies hat auch damit zu tun, dass eine subjektive Verzerrung des Gemessenen sehr gering ist, und die sensorgenerierten Daten damit vergleichsweise valide sind. Am Beispiel der Verkehrslenkung können Energieverbrauch, städtische Flächennutzung, Logistik und private Mobilität optimiert werden. Quartiers- oder Stadtteilbeobachtung können in Echtzeit erfolgen und städtische Ressourcen effizienter erfasst und eingesetzt werden. Die Analyse und Nutzung von Echtzeit-Sensordaten erfordert jedoch umfassende Kompetenzen und Ressourcen.

Bezug zu Smart City

Die Diskussion der Nutzung von sensorgenerierten Daten weist damit enge Bezüge zum Konzept der Smart Cities auf. Die technologische Sicht zu Smart City geht von der vollumfänglichen echtzeitdaten-vernetzten Steuerung aller Lebensäußerungen im Stadtraum aus. Die angestrebte Effizienzsteigerung der städtischen Systeme und Infrastrukturen soll gemäß BBSR (2015) jedoch nicht nur zu einer besseren CO_2-Bilanz und wirtschaftlichem Wachstum der Städte beitragen. Sie soll auch die Lebensqualität der Bürger verbessern, indem sie sich an deren Bedürfnissen und Verhaltensweisen orientiert, also „responsive" ist (Goldsmith, Crawford 2014).

Technologischer Zugang

Sensorgenerierte Daten weisen einen stark technologiegetriebenen Zugang zur Stadtentwicklung auf. Dies birgt aber auch die Gefahr einer

stark technologie- / datenorientierten Stadtentwicklung, welche ein echtzeitdaten-getriebenen Steuerungs- und Management-Verständnis stärker gewichtet als eine mittel- und längerfristige Planung (Bebauungspläne, Strategiepläne) (Batty 2013).

6.4 Nutzungspotenzial von Crowd Data

Crowdsourcing in der öffentlichen Verwaltung

Für eine Stadtverwaltung bietet sich, wie für jede größere Organisation, die Möglichkeit, ihre Vielzahl von sektoral und thematisch selbst generierten Daten für eine effektivere und effizientere Steuerung der Stadt in Wert zu setzen. Daneben benutzt die öffentliche Verwaltungen Crowdsourcing bereits heute für verschiedene Aufgaben. Crowdsourcing ist im Bereich des sogenannten Anliegenmanagements bereits so weit etabliert, dass auch standardisierte Lösungen auf dem Markt verfügbar sind[13]. Beim Anliegenmanagement werden Meldungen der Bürger*innen etwa zur öffentlichen Infrastruktur aufgenommen[14]. Weitere Anwendungen bestehen bei der Datensammlung und Informationsbeschaffung im Kultur- und Bildungsbereich (z. B. Katalogisierung einer undokumentierten Luftbildsammlung der Bibliothek der Eidgenössischen Technischen Hochschule (ETH) Zürich durch Crowdsourcing) oder im Krisenmanagement, z. B. in den Projekten Ushahidi (2017) und Missing Maps (2017) (vgl. auch OpenStreetMap Wiki, 2017). Weitere Beispiele für Crowdsourcing finden sich in Anhang A3.

Beispiel: Ordnungsamt Online

Seit dem Sommer 2016 können mit der App "Ordnungsamt Online" Missstände in der Nachbarschaft nun auch von unterwegs leichter und komfortabler gemeldet werden[15]. Kritische Stimmen weisen etwa auf fehlende Ressourcen zur Bearbeitung oder datenschutzrechtliche Probleme hin.[16]

Citizen-Science Ansätze

Anspruchsvoller, weil eine Vorbildung der jeweiligen Nutzergemeinschaft vorausgesetzt wird, sind Citizen Science-Ansätze, die bei Stadt- und Freiraumplanung oder Umweltfragen eingesetzt werden können (vgl. Kapitel 3.2; ein Beispiel eines Citizen Science-Projekts ist in Anhang A3 enthalten).

ePartizipation

Crowdsourcing weist auch große Schnittstellen auf zu Themen wie Partizipation oder eGovernment. Die Stadt Zürich in einer Auswertung von

[13] Z. B. https://www.markaspot.de/

[14] Z. B. der Schlaglochmelder des Bezirksamtes Steglitz-Zehlendorf (https://www.berlin.de/ba-steglitz-zehlendorf/politik-und-verwaltung/aemter/strassen-und-gruenflaechenamt/tiefbau/formular.87799.php)

[15] https://www.berlin.de/ordnungsamt-online/mobile-app/

[16] http://www.rbb-online.de/politik/beitrag/2016/07/berlin-bezirke-ordnungsaemter-setzen-auf-neue-app.html

Praxisbeispielen von ePartizipation zum Fazit: „ePartizipation ist kein Wundermittel, aber mit den zurzeit raschen Veränderungen der Technologie dürften sich auch die Möglichkeiten für ePartizipation künftig rasch erweitern. Es gilt insbesondere zu beobachten, ob und wie sich die Möglichkeiten der elektronischen Textauswertung verbessern und welche Arten nicht-textorientierter Tools wie Spiele und Simulationen im Bereich ePartizipation entwickelt werden und diesen bereichern können (Stadt Zürich 2013). Haan und Höffken (2015) konstatieren, dass das technologische Potenzial mit dem Trend des Selbstmachens einhergeht. Aus dieser Verbindung entstehe ein spannendes Themenfeld, das innovative Ansätze und Methoden für Mitgestaltung und Teilhabe im urbanen Diskurs ermöglicht und perspektivisch neue Entwicklungsoptionen für Stadtplanung und Verwaltungshandeln bietet.

Beispiel: Bürger baut Stadt

Das Portal „Bürger baut Stadt" schafft einen Überblick zu geplanten Bauvorhaben in Berlin. Es bereitet die Daten der Ratsinformationssysteme neu auf, ergänzt diese durch Daten aus dem Amtsblatt Berlin sowie offene Datensätze und stellt sie auf einer digitalen Karte dar.

Abbildung 13: www.buergerbautstadt.de

Beispiel: Leerstandsmelder

Auf der Website leerstandsmelder.de werden auf einer interaktiven Karte als leerstehende Wohnungen, Gebäude oder Gewerbeflächen dargestellt. Eine leerstehende Wohnung melden kann generell jeder, der sich auf der Leerstandsmelder-Seite anmeldet. Die Leerstandsmelder-Seite ist in verschiedenen deutschen Städten präsent. In Berlin liegt die Betreuung beim Verein openberlin.

Crowdsourcing mit Legitimationsfrage

Bei allen Ansätzen des Crowdsourcing stellt sich neben der grundlegenden Frage des Datenschutzes die Legitimationsfrage: Wie weit kann und soll die hoheitliche Verwaltung Aufgaben an steuerzahlende Subjekte – Bürger*innen – abgeben, ohne ihre implizite Begründungshoheit staatlichen Handelns zu untergraben oder gar zu verlieren. Kritisch, zumindest

diskussionswürdig wird es, wenn die öffentliche Hand aus ihrer Finanzknappheit eine Tugend macht und die Unterhalts-Aufgabe an ihre Bürger delegiert, zumindest die dezentrale, systematische Erkennung. Anderseits könnten diese Informationen gekoppelt mit Sensordaten gestützten Systemen das Predictive Maintenance für Infrastrukturen aller Art unterstützen.

Die Durchführung eigener Crowdsourcing-Projekte bietet den Verwaltungen dabei verschiedene Chancen und Risiken (Straumann, 2012).[17]

Tabelle 7: Chancen und Risiken von Crowdsourcing

Chancen	Risiken
Hohe Datenaktualität	Genauigkeitsprobleme
Günstige Erhebung	Aufwand zur Bewirtschaftung und Pflege der Community
Neue Gesprächspartner*innen und Datenlieferant*innen	Verzerrungen in der Nutzer*innenstruktur (Selbstselektion der Teilnehmenden)
Bürgernähe, Instrument zur Öffentlichkeitsbeteiligung	Wecken von vielfältigen Erwartungen, z. B. bezüglich Aufnahme und Umsetzung der persönlichen Wünsche
Spezifische Ausrichtung auf eine definierte Thematik und Fragestellung (z. B. im Gegensatz zu Big Data-Anwendungen, wo die Datenerfassung fremddefiniert ist)	Risiko eines in der Öffentlichkeit direkt sichtbaren Fehlschlags (z. B. Nicht-Aufbringen der kritischen Masse an Teilnehmenden, Versagen des Community Managements bspw. „Datenvandalismus“)
Positives Image	Beeinflussbarkeit durch Falschinformation oder gezielte Manipulation

Zunehmende Möglichkeiten von Crowdsourcing

Der steigende Grad der digitalen Vernetzung, die ständige Weiterentwicklung der technischen Möglichkeiten und die zunehmend einfacher werdende Handhabung von mobilen Angeboten werden die Möglichkeiten des Crowdsourcings weiter ausdehnen. Höffken (2014) bezeichnet die mobilen Geräte als dezentrale städtische Infrastruktur, die gleichzeitig Kommunikations- und Produktionsmittel, Analyse-, Erfassungs-, Visualisierungs- und Organisationstool sind. Damit steigen auch die Möglichkeiten zur Nutzung solcher Informationen für die Stadtentwicklungsplanung, beispielsweise im Hinblick auf eine nachhaltige Mobilität (Attard et al. 2016).

[17] Zu den generellen Chancen und Risiken von Big Data und Crowd Data siehe Kap. 8.1.

7 Nutzungspotenzial für die Stadtentwicklungspläne (StEP)

In Kapitel 7 werden die drei in Überprüfung bzw. Fortschreibung befindlichen Stadtentwicklungspläne (StEP) „Zentren", „Wohnen" sowie „Industrie und Gewerbe" näher betrachtet. Die gemachten Aussagen können als Beitrag zur künftigen Aktualisierung der StEPs dienen. Basierend auf den generellen Einschätzungen aus Kapitel 6 werden Big Data Quellen und deren Aussagekraft zu den Themen der einzelnen StEPs aufgezeigt.

Ratingplattformen, Immobilien-, Unterkunfts- und Mobilitätsportale besitzen für die drei betrachteten Stadtentwicklungspläne die höchste Relevanz und eine potenzielle Aussagekraft. Die Daten vermitteln dabei als Momentaufnahme einen Eindruck eines Sachverhaltes und liefern Hinweise auf näher zu betrachtende Zusammenhänge oder Fragestellungen. Dies insbesondere in Kombination mit weiteren „konventionellen" Daten. Spezifische Anwendungen lassen sich angesichts der erkennbaren Einschränkungen noch nicht benennen. Konkrete Nutzungsmöglichkeiten für ausgewählte Datensätze wären weiter zu prüfen.

7.1 StEP Zentren 3

7.1.1 Beschreibung

Inhalte

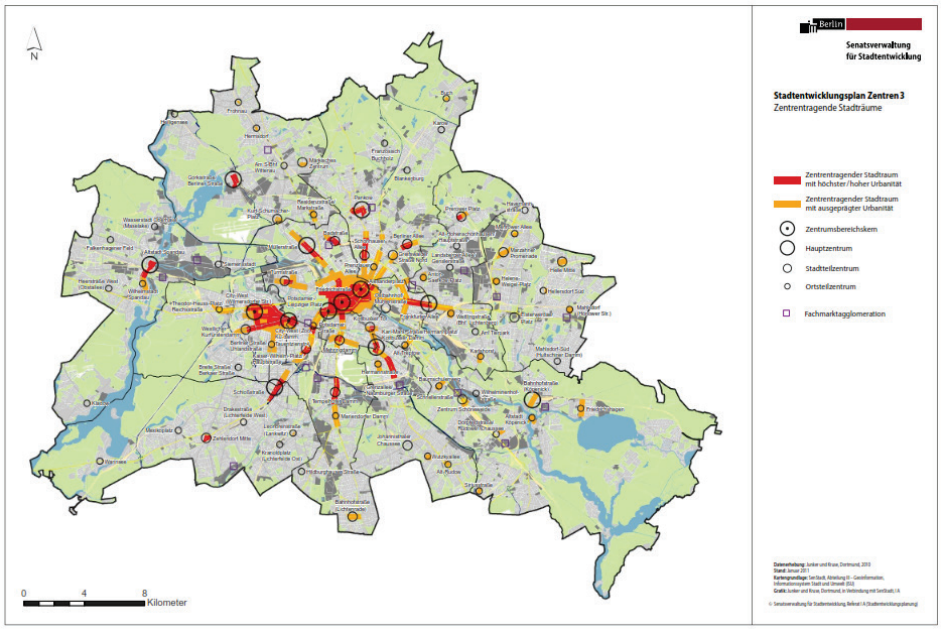
Der StEP Zentren 3 wurde 2011 beschlossen. Er enthält folgende Elemente:

- Sechs Ziele als inhaltliche Schwerpunkte
- Zehn Leitlinien, die verdeutlichen, wie die Ziele umgesetzt werden sollen
- Sechs Steuerungsgrundsätze, welche die Umsetzung konkretisieren

Zentrenkonzept und Ausstattungsmerkmale

Grundlage des Stadtentwicklungsplans bildet das hierarchisch gegliederte, polyzentrale Zentrensystem. Für die einzelnen Zentrentypen innerhalb der Hierarchie sind Ausstattungsqualitäten in idealtypischer Form beschrieben.

Abbildung 14: StEP Zentren, Zentrenhierarchie und zentrentragende Stadträume



Themen StEP Zentren

Abbildung 16 benennt diejenigen Themen, gestützt auf die Ziele und Inhalte des StEP Zentren 3, welche hinsichtlich der Aussagekraft von Big Data Quellen geprüft werden.

Abbildung 15: Themen StEP Zentren

7.1.2 Datenquellen mit Aussagekraft zumStEP Zentren

Kategorien von Datenquellen mit Bezug zum StEP Zentren

Abbildung 17 zeigt diejenigen Kategorien von Datenquellen, die zumindest eine geringe Aussagekraft hinsichtlich eines oder mehrerer Themen des StEP aufweisen. Für eine differenzierte Betrachtung wurde die Aussagekraft auf einer Skala von 0 (ohne thematischen Bezug) bis 3 (aus den Daten lassen sich unmittelbare und umfassende Informationen zum Thema ableiten) beurteilt.

Aussagen zur (touristischen) Nachfrage und zur Zentrumsausstattung

Es sind verschiedene Datenquellen vorhanden, welche Aussagen zu den Themen des StEP Zentren machen können. Diese betrifft vor allem Informationen zu (touristischer) Nachfrage und Zentrumsausstattung.

Foto- und Ratingplattformen

Gerade Besucher*innen nutzen vielfach soziale Medien, etwa auf Foto- oder Ratingplattformen. Auf Plattformen wie Tripadvisor beurteilen Nutzer*innen direkt Angebote und lassen damit Präferenzen der (touristischen) Nachfrage erkennbar werden. Auch über Google Maps besteht die Möglichkeit, Rezensionen zu jeglichen Geschäften oder Einrichtungen zu verfassen, welche dort verzeichnet sind. Neben Informationen wie Adresse, Kontakt oder Öffnungszeiten wird auch das zeitliche Verlaufsmuster der Besuche an diesem Ort dargestellt. Zur Ermittlung dieser Informationen werden Daten genutzt, die von den Nutzenden stammen. Stoßzeiten und Besuchsdauer werden mithilfe von aggregierten und anonymisierten Daten von Nutzer*innen berechnet, die den Google Standortverlauf aktiviert haben.

Abbildung 16: Kategorien von Datenquellen mit Aussagekraft zum StEP Zentren

geringe Aussagekraft
mittlere Aussagekraft
hohe Aussagekraft

Kategorie der Plattform	*Themen StEP Zentren*	
Nutzergenerierte Daten		
Fotoplattformen	Zentrumsausstattung, touristische Nachfrage	mittlere Aussagekraft
Ratingplattformen	Touristische Nachfrage, Nachfrage, Zentrumsausstattung	hohe Aussagekraft
Kartenportale	Touristische Nachfrage, Nachfrage, Zentrumsausstattung	mittlere Aussagekraft
Social Media	Touristische Nachfrage, Nachfrage, Zentrumsausstattung	geringe Aussagekraft
Suchportale	Touristische Nachfrage, Nachfrage	geringe Aussagekraft
Businessnetzwerke	Nachfrage	geringe Aussagekraft
Transaktionsgenerierte Daten		
Mobilitätsplattformen	Einzugsbereiche, Nachfrage	mittlere Aussagekraft
Unterkünfte	Touristische Nachfrage	mittlere Aussagekraft
Tauschplattformen	Nachfrage	geringe Aussagekraft
Online-Handel	Nachfrage	geringe Aussagekraft
Mobilitätsdaten	Zentrumsausstattung, Einzugsbereiche	geringe Aussagekraft
Sensorgenerierte Daten		
Navigationssysteme	Einzugsbereiche	geringe Aussagekraft
Mobilfunkdaten	Städtebauliche Situation, Einzugsbereiche, Nachfrage, Touristische Nachfrage	mittlere Aussagekraft
Überwachungskamerasysteme (CCTV)	Städtebauliche Situation, Touristische Nachfrage	geringe Aussagekraft

Rückschlüsse aus Verhaltensmustern

Aus den Verhaltensmustern lassen sich Rückschlüsse auf die Zentrumsausstattung und insbesondere auch deren Attraktivität bzw. Qualität ableiten. Diese bedürfen jedoch einer zusätzlichen Interpretation bzw. einer Erklärung der Nutzerhandlungen. So kann beispielsweise das Hochladen eines Fotos aus ganz unterschiedlichen Motiven erfolgen.

Auch das GPS-Tracking oder die Rekonstruktion von Trajektorien beispielsweise aus Daten von Fotosharingplattformen zeigen spezifisches Verhalten. Aus diesem lassen sich jedoch nur mit erheblichen Unsicherheiten raumbezogene Aussagen etwa zur Umfeldqualität herauslesen (z. B. Attraktivität von Wegen / Verbindungen). Auch aus den Aufenthaltsdichten und Bewegungsmustern von Mobilfunkdaten können Rückschlüsse auf Nachfrager bzw. die Einzugsbereiche von Zentren gewonnen werden.

Weitere Sondierung notwendig

Es wäre weiter zu sondieren, ob und wenn ja, welche Analysen im Zusammenhang mit dem StEP Zentren durch die Auswertung von Big Data und Crowd Data untersetzt werden könnten, um weitergehenden Planungsansätzen zusätzlich erforderliche Datengrundlagen zur Verfügung zu stellen.

7.2 StEP Wohnen

7.2.1 Beschreibung

Inhalte

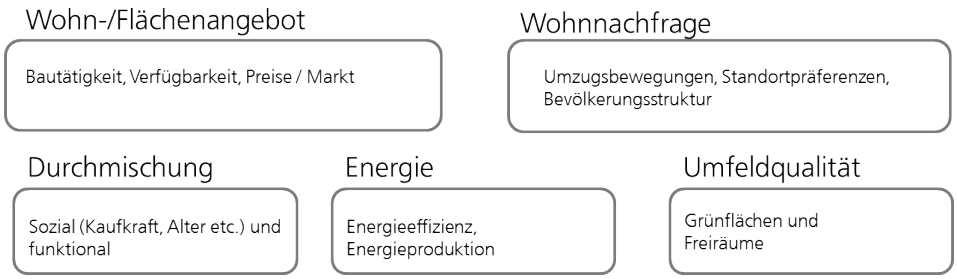
Die Beschlussfassung zum StEP Wohnen 2025 erfolgte im Jahr 2014. Er enthält folgende Elemente:

- Abschätzungen zur zukünftigen Entwicklung von Nachfrage und Wohnbauflächenpotenzialen
- Sieben Leitlinien, mit denen zentrale thematische stadtentwicklungs- und wohnungspolitische Ziele verfolgt werden. Für jede der Leitlinien ist ein Ziel sowie Mittel und Wege zu dessen Erreichung formuliert.
- Maßnahmenräume
- 15 prioritäre Aufgaben

Themen StEP Wohnen

Abbildung 18 benennt diejenigen Themen, gestützt auf die Ziele und Inhalte des StEP Wohnen, welche hinsichtlich der Aussagekraft von Big Data Quellen geprüft werden.

Abbildung 17: Themen des StEP Wohnen



7.2.2 Datenquellen mit Aussagekraft zum StEP Wohnen

Kategorien von Datenquellen mit Bezug zum StEP Wohnen

Abbildung 19 zeigt diejenigen Kategorien von Datenquellen, die zumindest eine geringe Aussagekraft hinsichtlich eines oder mehrerer Themen des StEP aufweisen.

Immobilienportale mit direktem Bezug

Mit umfangreichen Informationen zum Wohn- und Flächenangebot sowie zur Wohnnachfrage lassen sich aus Immobilienportalen relevante Aussagen ableiten. Auch das Suchverhalten bzw. Suchattribute auf Immobilienplattformen können anzeigen, welche Nachfrage-Merkmale sich verändern. Entsprechende Informationen werden bereits heute vielerorts verwendet.

Angebots- und Nachfrageverhalten

Weitere Sondierung notwendig

Daten von Immobilienportalen werden bereits heute für die Wohnungsmarktberichterstattung genutzt[18]. Es wäre weiter zu sondieren, ob eine darüberhinausgehende Nutzung von Big Data und Crowd Data im Zusammenhang mit dem StEP Wohnen einen weitergehenden Erkenntnisgewinn liefert.

Abbildung 18: Kategorien von Datenquellen mit Aussagekraft zum StEP Wohnen

geringe Aussagekraft
mittlere Aussagekraft
hohe Aussagekraft

Kategorie der Plattform	*Themen StEP Wohnen*	
Nutzergenerierte Daten		
Fotoplattformen	Lebensqualität	geringe Aussagekraft
Ratingplattformen	Umfeldqualität	geringe Aussagekraft
Kartenportale	Umfeldqualität	geringe Aussagekraft
GPS Tracking	Umfeldqualität	mittlere Aussagekraft
Datingportale	Durchmischung	geringe Aussagekraft
Transaktionsgenerierte Daten		
Mobilitätsplattformen	Durchmischung	geringe Aussagekraft
Immobilienportale	Wohn-/Flächenangebot, Wohnnachfrage	hohe Aussagekraft
Unterkünfte	Wohnangebot	mittlere Aussagekraft
Online-Handel	Durchmischung	geringe Aussagekraft
Mobilitätsdaten	Umfeldqualität	geringe Aussagekraft
Sensorgenerierte Daten		
Überwachungskamerasysteme (CCTV)	Umfeldqualität	geringe Aussagekraft
Gebäudedaten	Energie, Umfeldqualität	mittlere Aussagekraft

Beurteilung der Umfeldqualitäten

Im Weiteren können etwa Daten von GPS Tracking Systemen, mit der Abbildung von räumlichen Verhaltensmustern, zur Beurteilung der Umfeldqualität genutzt werden. Aber auch hier bleibt die Interpretation der Motivlagen schwierig. Auch aus Gebäude- und Wohnungsdaten lassen sich Aussagen ableiten. Weitere Datenquellen weisen Bezüge zu Um-

[18] IBB Wohnungsmarktbericht 2016

feldqualität und Durchmischung auf, dies aber nur mit geringer Aussagekraft.

7.3 StEP Industrie und Gewerbe

7.3.1 Beschreibung

Inhalte

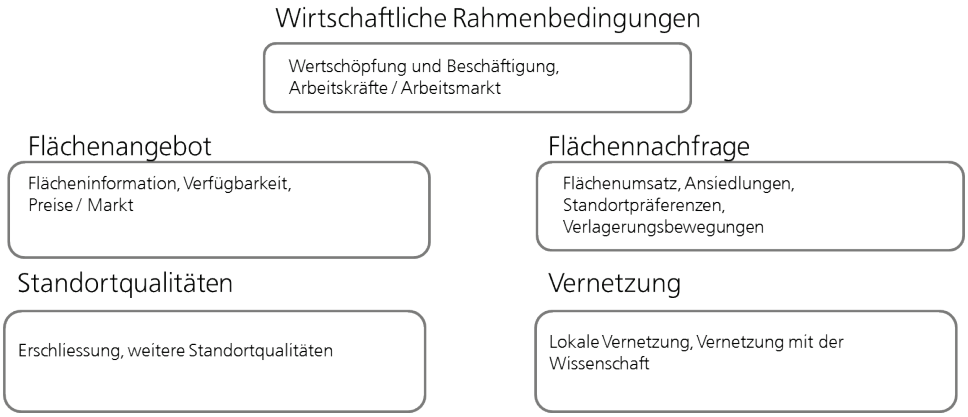
Die Beschlussfassung zum StEP Industrie und Gewerbe erfolgte im Jahr 2011. Er enthält folgende Elemente:

- Analysen von Flächenpotenzialen
- Zehn Leitlinien
- Räumliches Modell und Konzept
- Entwicklungskonzept für den produktionsgeprägten Bereich (EpB)

Themen StEP Industrie und Gewerbe

Abbildung 20 benennt diejenigen Themen, gestützt auf die Ziele und Inhalte des StEP Industrie und Gewerbe, welche hinsichtlich der Aussagekraft von Big Data Quellen geprüft werden.

Abbildung 19: Themen des StEP Industrie und Gewerbe



7.3.2 Datenquellen mit Aussagekraft zum StEP Industrie und Gewerbe

Kategorien von Datenquellen mit Bezug zum StEP Industrie und Gewerbe

Abbildung 21 zeigt diejenigen Kategorien von Datenquellen, die zumindest eine geringe Aussagekraft hinsichtlich eines oder mehrerer Themen des StEP aufweisen.

Jobbörsen

Im Themenbereich StEP Industrie und Gewerbe bieten vor allem Jobbörsen Hinweise auf wirtschaftliche Rahmenbedingungen (Arbeitsplatzentwicklung nach Branche). Stellenangebote widerspiegeln die Zukunftserwartungen der Unternehmen und mittelfristig relevante strukturelle Trends in der Arbeitskräftenachfrage.

Abbildung 20: Kategorien von Datenquellen mit Aussagekraft zum StEP Industrie und Gewerbe

geringe Aussagekraft
mittlere Aussagekraft
hohe Aussagekraft

Kategorie der Plattform	*Themen StEP Industrie und Gewerbe*	
Nutzergenerierte Daten		
Social Media	Wirtschaftliche Rahmenbedingungen	geringe Aussagekraft
Suchportale	Wirtschaftliche Rahmenbedingungen	geringe Aussagekraft
Businessnetzwerke	Wirtschaftliche Rahmenbedingungen, Vernetzung	mittlere Aussagekraft
Transaktionsgenerierte Daten		
Mobilitätsplattformen	Erschnliessung	geringe Aussagekraft
Immobilienportale	Flächenangebot, Flächennachfrage	hohe Aussagekraft
Ausschreibungsplattformen	Wirtschaftliche Rahmenbedingungen	geringe Aussagekraft
Jobbörsen	Wirtschaftliche Rahmenbedingungen	hohe Aussagekraft
Mobilitätsdaten	Standortqualitäten	geringe Aussagekraft
Sensorgenerierte Daten		
Navigationssysteme	Standortqualitäten	geringe Aussagekraft
Gebäudedaten	Standortqualitäten	geringe Aussagekraft

Immobilienportale

Immobilienportale weisen mit Informationen zu Flächenangebot und Nachfrage im Bereich Gewerbe und Industrie eine hohe Aussagekraft auf.

Businessnetzwerke

Businessnetzwerke können Vernetzungen aufzeigen. Die Verfügbarkeit und Nutzbarkeit konkreter Informationen wäre jedoch noch vertieft zu prüfen.

Nutzungspotenzial noch offen

Kernanliegen des StEP Industrie und Gewerbe ist die strategische Flächenvorsorge für Berlins wachsende Wirtschaft. Die hier skizzierten Anwendungsoptionen lassen offen, ob und mit welchem Aufwand die Ziele des StEP Industrie und Gewerbe als räumlich wirksamem Plan durch Big Data und Crowd Data Nutzungen sinnvoll untersetzt werden können.

8 Zusammenfassende Einschätzungen und Ausblick

Maßgeblich gestützt auf die Diskussionen in einem Fachgespräch werden in Kapitel 8 Chancen und Risiken im Umgang mit Big Data und Crowd Data erörtert. Basierend auf den verschiedenen Analyseschritten werden letztlich unterschiedliche Ansatzpunkte und konkrete Maßnahmen aufgezeigt, wie sich die Stadtentwicklungsplanung vertieft mit Big Data und Crowd Data auseinandersetzen und diese verstärkt für ihre Zwecke nutzbar machen kann.

Die Nutzung von Big Data und Crowd Data bietet das Potenzial einer schnelleren und kostengünstigeren Verfügbarkeit von zusätzlichen Daten und Informationen. Neuartige bzw. verbesserte Erkenntnisse sollen letztlich zu einer bedarfsgerechteren Planung beitragen. Die grundsätzlichen Potenziale und Chancen werden jedoch (noch) durch verschiedene „Aber" relativiert. Angesichts der rasanten Entwicklung der Thematik sind Fragen hinsichtlich der konkreten Nutzungsmöglichkeiten, der dazu erforderlichen Kompetenzen und Ressourcen innerhalb der Verwaltung, zu Datenschutz und Privatsphäre sowie bezüglich (planungs-)rechtlicher Einschränkungen noch weitgehend offen. Es ist davon auszugehen, dass sich Big Data rasant weiter entwickeln wird. Entsprechend erscheint eine vertiefte Auseinandersetzung mit Big Data und Crowd Data lohnenswert.

8.1 Chancen und Risiken im Umgang mit Big Data und Crowd Data – Erkenntnisse des Fachgesprächs

Eine abschließende Beurteilung der Chancen und Risiken im Umgang mit Big Data und Crowd Data ist im Rahmen dieses Projektes angesichts der Breite des Themas und der Verschiedenartigkeit der Datenquellen, Fragestellungen und denkbaren Anwendungen nicht möglich. Die folgenden Einschätzungen basieren maßgeblich auf den Diskussionen des Fachgesprächs am 17. Februar 2017 (siehe Anhang A7). Bei der Betrachtung der Chancen und Risiken gilt es zu beachten, dass die verschiedenen Typen von Big Data sowie Crowd Data ganz unterschiedliche Charakteristika und damit auch spezifische Chancen und Risiken aufweisen. Eine entsprechende Differenzierung kann nachfolgend nur ansatzweise vorgenommen werden.

8.1.1 Chancen im Umgang mit Big Data und Crowd Data

Neue Daten ermöglichen neue Einblicke

Mit Big Data entsteht in der Regel eine große Menge an Daten. Mit Big Data ist für die Teilnehmer*innen des Fachgesprächs aber auch die Erwartung verbunden, dass aus den Big Data-Quellen „bessere" Daten und Informationen resultieren. „Besser" umfasst in diesem Zusammenhang verschiedenen Dimensionen: Zum einen können dies neue Daten sein, welche bislang nicht erhoben werden konnten. Es wird erwartet, dass damit neuartige Zusammenhänge erkannt und Verbindungen hergestellt werden können. Mit neuen Daten lassen sich aus Sicht der Stadtentwicklungsplanung auch andere Bilder der Stadt generieren, die unbekannte oder noch wenig bekannte Facetten aufzeigen.

Räumliche Präzision und erhöhte Aktualität

Eine zweite Dimension umfasst die räumliche Präzision. Aus einzelnen Quellen lassen sich räumlich hoch aufgelöste und feingliedrige Daten gewinnen, welche Rückschlüsse und Aussagen auf einer sehr kleinräumigen Ebene ermöglichen können. Als wohl bedeutendste Verbesserung wird die in der Regel hohe Aktualität der Daten benannt. Die zeitnahe Verfügbarkeit, bis hin zu Echtzeit-Daten, biete die Chance, Entwicklungen schneller zu erkennen und auch entsprechend kurzfristiger darauf reagieren zu können.

Reduzierung des Aufwands

Eine weitere Chance wird von den Teilnehmer*innen in einem geringeren Aufwand für die Datenerhebung und -beschaffung ausgemacht. Big Data und Crowd Data kann die Möglichkeit bieten, die Bürger*innen von klassischen Erhebungen zu entlasten und gleichzeitig den administrativen Aufwand auf Seiten der öffentlichen Hand zu reduzieren. Zudem bieten neue Datenquellen auch die Möglichkeit, die nicht unerhebliche *non-response*-Rate klassischer Erhebungs- und Befragungsinstrumente zu reduzieren.

Stärkung der Nähe zu Bürger*innen

Eine schnellere Verfügbarkeit neuer Daten kann die Bürger*innen-Nähe stärken, indem Entwicklungstrends und Fragestellungen zeitnah aufgegriffen und analysiert werden können. Verfahren und Prozesse könnten vereinfacht und beschleunigt werden. Mit außerhalb von strukturierten, amtlichen Erhebungsverfahren gesammelten Daten kann die „Lebenswirklichkeit" der Bürger*innen auf andere Weise, möglicherweise adäquater und präziser beschrieben werden. Ein großes Potenzial wird in dieser Hinsicht insbesondere dem Crowdsourcing attestiert[19]. Es schafft eine neue Möglichkeit, Bedürfnisse der Bürger*innen zu erkennen, den Dialog auf neuen Kanälen verstärken zu können und damit eine partizi-

[19] Siehe dazu Tabelle 7: Chancen und Risiken von Crowdsourcing

pativere, offenere und bürger*innenorientierte Stadtentwicklung zu konzipieren.

Bedarfsgerechtere Planung

Neue und bessere Daten können in der Einschätzung der Teilnehmer*innen des Fachgesprächs kein Selbstzweck sein. Letztlich muss es aus Sicht der Stadtentwicklungsplanung darum gehen, bedarfsgerechtere Entscheidungen treffen und Planungen vornehmen zu können. Eine Benennung von Möglichkeiten, wie Big Data und Crowd Data dabei ganz konkret genutzt werden können, fällt im heutigen Wissensstand allerdings noch schwer. Aus diesem Grund soll der hier vorliegende Bericht den Wissenstand erhöhen.

Weitere Entwicklung wird die Chancen weiter erhöhen

Übereinstimmend gehen die Teilnehmenden des Fachgesprächs davon aus, dass die Möglichkeiten und Chancen von Big Data heute erst ansatzweise erkennbar sind. Big Data entwickeln sich rasant weiter, wachsen enorm rasch und werden zunehmend an Bedeutung gewinnen. Gleichzeitig dürften sich auch die Analysemethoden und das Know-how in der Datenauswertung weiterentwickeln. Diese wiederum werden längerfristig Einzug halten in nutzer*innenfreundliche breit verfügbare Software und so für einen deutlich größeren Personenkreis nutzbar sein als bisher. Mit verbesserten Analysemöglichkeiten lassen sich aus den Daten auch mehr Informationen gewinnen und – so die Hoffnung – die oben genannten Chancen erhöhen und greifbarer machen.

8.1.2 Risiken im Umgang mit Big Data und Crowd Data

Chancen ja, aber …

Die grundsätzlichen Potenziale und Chancen, wie sie die Teilnehmenden des Fachgesprächs Big Data und Crowd Data beimessen, werden durch verschiedene „Aber“ ", d. h. kritische Faktoren, relativiert.

Nutzungsmöglichkeiten kurzfristig überschätzt

Big Data hat in den letzten Jahren einen wahren Hype erlebt.[20] Es besteht, so die Einschätzung der Teilnehmenden, die Gefahr, dass die Nutzungsmöglichkeiten von Big Data – wie bei vielen Technologien – kurzfristig überschätzt werden. Gleichzeitig wird aber auch festgehalten, dass die langfristigen Auswirkungen in vielen Fällen unterschätzt werden.

Eingeschränkte Datenqualität begünstigt Fehlinterpretationen

Ein wesentlicher Vorbehalt wird hinsichtlich der Datenqualität angemeldet. Diese Einschränkung betrifft insbesondere nutzergenerierte Daten. Wie in Kapitel 6.1 dargelegt, ist Ihre Qualität vielfach beschränkt oder lässt sich – im noch schlechteren Fall – nur begrenzt oder kaum beurteilen. Zudem macht die hohe Volatilität der Plattformen mit schnell wechselnden Funktionalitäten und wechselndem Markterfolg auch Vergleiche über mittel- und langfristige Zeiträume für manche Fragestellungen fast

[20] https://www.datanami.com/2015/08/26/why-gartner-dropped-big-data-off-the-hype-curve/

unmöglich. Die Daten stellen deshalb häufig nur Momentaufnahmen dar, deren Interpretation im Rahmen längerfristiger Veränderungen zumindest schwierig ist. Deshalb besteht die Gefahr, dass Entwicklungen falsch interpretiert oder kurzfristige Veränderungen überbewertet werden, woraus Trugschlüsse resultieren.

Zusätzliche Kompetenzen im Umgang mit Daten erforderlich

Der Umgang mit Big Data und Crowd Data stellt gegenüber der Sammlung, Verarbeitung und Auswertung traditioneller Daten erhöhte Anforderungen an die öffentliche Hand. Dies betrifft einerseits die technologische Ausstattung zur Verarbeitung von Big Data und auch besonders von Echtzeitdaten(strömen). Noch stärker ins Gewicht fällt jedoch, dass auch neue Kompetenzen im Datenhandling und in der Datenanalyse notwendig sind. Dieses Know-how ist bislang verwaltungsintern noch kaum vorhanden. In die Zukunft gerichtet, stellt sich die Frage, wie weit die entsprechenden Kompetenzen innerhalb der Verwaltung vorhanden sein bzw. aufgebaut werden sollen und können und in welchem Ausmaß ein Bezug von externem Wissen notwendig bzw. zweckmäßiger ist.

Fehlende Ressourcen

In der Diskussion im Fachgespräch wurde die Frage der Ressourcenverfügbarkeit aufgeworfen. Es wurde die Befürchtung geäußert, dass die Potenziale zur Kostenreduktion bei der Datenerhebung durch den aus der Datenverarbeitung resultierenden Aufwand wieder zunichtegemacht oder gar ins Gegenteil verkehrt würden.

Vor lauter Bäumen den Wald nicht mehr sehen

Eine steigende Datenmenge kann Entscheidungen auch erschweren, so ein weiteres Fazit aus dem Fachgespräch. Gelingt es nicht, die Daten in überschaubare und verständliche Informationen zu verdichten, besteht die Gefahr, in Daten zu „versinken". Fehlender Überblick, die erhöhte Komplexität und daraus resultierende widersprüchliche Interpretationen würden den intendierten Erkenntnisgewinn erschweren und dadurch letztlich Entscheidungsprozesse verzögern oder auch in eine falsche Richtung lenken.

Einbettung in den gesellschaftlichen Diskurs

Die Betrachtung der Nutzungsmöglichkeiten von Big Data und Crowd Data in der Stadtentwicklungsplanung steht im Kontext eines größeren gesellschaftlichen Diskurses. Die Debatte um digitale Grundrechte wird europaweit geführt.[21] In der öffentlichen Debatte stehen hohen Erwartungen bezüglich Nutzungsmöglichkeiten häufig auch tiefgreifende Befürchtungen und Ängste gegenüber, welche in Big Data eine fundamentale Bedrohung für Freiheit und Privatsphäre von Bürger*innen sehen. Die Nutzung von Big Data birgt auch ethische Herausforderungen (SATW 2017). Eine skeptische und ablehnende Haltung gegenüber einem

[21] Siehe etwa den Vorschlag einer Gruppe von Bürgerinnen und Bürgern für eine „Charta der Digitalen Grundrechte in der Europäischen Union", der am 5.12.2016 dem Europäischen Parlament in Brüssel und der Öffentlichkeit zur weiteren Diskussion übergeben wurde (https://digitalcharta.eu/).

überwachenden und datensammelnden Staat kann sich ganz konkret in einer Haltung äußern, welche neuartige Datenerhebungen oder -nutzungen im Rahmen der Stadtentwicklungsplanung verweigert.

Kritischer Umgang mit nicht-hoheitlichen Daten

Der Umgang mit nicht-hoheitlichen Daten birgt für die öffentliche Hand zahlreiche Herausforderungen. Der Bezug und die Nutzung von privaten Daten und die Kooperation mit privaten Anbietern werfen zahlreiche rechtliche Fragen auf. Insbesondere natürlich, wenn private Daten in offizielle Statistiksysteme integriert werden (siehe z. B. UNECE 2016) oder als formale Planungsgrundlagen dienen sollen. Die Konzepte der Datenvermeidung und der Datensparsamkeit als Teil der datenschutzrechtlichen Bestimmungen sehen vor, dass nur so viele Daten gesammelt werden, wie für die jeweilige, i. d. R. vorab definierten Anwendung unbedingt notwendig sind. Dies erschwert eine explorative Datenanalyse und Datennutzung, wie sie für den Umgang mit Big Data häufig charakteristisch sind.

8.2 Ansatzpunkte zur vertieften Auseinandersetzung mit Big Data und Crowd Data

Geringes Wissen und wenig Erfahrung

Die in Kapitel 8.1 allgemein formulierten Chancen und Risiken resultieren einerseits aus der Breite des Datenangebots, der Vielzahl von Analysemöglichkeiten und denkbaren Anwendungen von Big Data und Crowd Data. Die thematische Breite lässt nur begrenzt für den gesamten Themenkreis gültige, konkrete Aussagen zu. Andererseits resultiert diese Unschärfe aber auch aus der Tatsache, dass heute in der Stadtentwicklungsplanung, und in der Verwaltung generell, aber auch in der Wirtschaft und der Zivilgesellschaft kaum fundiertes Wissen und nur wenige fundierte Erfahrungen im Umgang mit Big Data und Crowd Data vorhanden sind.

Vertiefte Auseinandersetzung mit Big Data und Crowd Data

Trotz verschiedener Risiken lassen es die heutigen und noch mehr die künftig erwarteten Chancen und Nutzungspotenziale als lohnend erscheinen, sich aus Sicht der Stadtentwicklungsplanung weiter und vertieft mit Big Data und Crowd Data auseinanderzusetzen. Dazu ist es sinnvoll und notwendig, auf verschiedenen Ebenen anzusetzen. In den Kapiteln 8.2.1 bis 8.2.5 sind, im Sinne einer Ideensammlung, fünf mögliche Ansatzpunkte benannt und konkrete Aktivitäten und Maßnahmen aufgeführt, die darauf abzielen, Big Data und Crowd Data für die Stadtentwicklungsplanung verstärkt nutzbar zu machen Teilweise sind es isolierte Maßnahmen, welche einzeln bzw. parallel umgesetzt werden können. Teilweise bauen die Aktivitäten aber auch inhaltlich aufeinander auf und erfordern ein schrittweises Vorgehen.

8.2.1 Eigene Kompetenzen aufbauen

Prozesse verstehen, um die richtigen Fragen zu stellen

Big Data durchdringt zunehmend mehr Bereiche des täglichen Lebens und damit auch der stadträumlichen Entwicklung. Eine Stadtentwicklungsplanung, die auf die daraus resultierenden Anforderungen Antworten formulieren will, wird nicht umhinkommen, ihre Kompetenzen im Umgang mit Big Data und Crowd Data zu stärken. Mit Big Data ist es notwendiger denn je, die richtigen Fragen zu stellen. Ansonsten besteht die Gefahr, sich im wachsenden „Datenberg" zu verlieren.

Folgende konkrete Schritte sind möglich:

- Erschließung der Erfahrungen jüngerer Mitarbeitender als Nutzer*innen von z. B. Apps, die Big Data oder Crowd Data generieren.
- Teilnahme an Big- und Crowd Data-bezogenen Tagungen und Veranstaltungen sowie Weitergabe und Reflektion des Gehörten in internen Diskussionen

- Durchführung von internen Fachdiskussionen und Weiterbildungsveranstaltungen mit externen Referent*innen
- Beteiligung an der Fachdiskussion auf nationaler Ebene (u. a. BBSR, z. B. im Rahmen des Forschungsprojektes „Smart Cities – Entwicklung eines stadtentwicklungspolitischen Handlungsrahmens“)
- Themenbezogene Kontakte mit anderen Großstädten, z. B. mit Wien
- Prüfung von Anstellungsmöglichkeiten für Mitarbeitende mit spezifischem (Daten-)Wissen

8.2.2 Den verwaltungsübergreifender Austausch verstärken

Gemeinsame Weiterentwicklung des Themas

Die Nutzung von Big Data und Crowd Data mag in der Stadtentwicklungsplanung teilweise themenspezifisch sein. Darüber hinaus stellen sich jedoch in allen Senatsverwaltungen vergleichbare technische, rechtliche oder organisatorische Fragen. Ein verstärkter verwaltungsübergreifender Austausch kann dabei helfen, Erfahrungen weiter zu tragen, im Dialog zu lernen und letztlich eine gemeinsame Haltung und Strategie im Umgang mit Big Data zu entwickeln.

Folgende konkrete Schritte sind möglich:

- Durchführung einer Umfrage bei den Senatsverwaltungen sowie den Bezirksämtern zur Nutzung von Big Data. Damit kann erfasst werden, mit welchen Daten zu welchem Zweck bereits gearbeitet wird und bei wem entsprechende Erfahrungen vorhanden sind.
- Identifikation von thematischen Schnittstellen und themenübergreifenden Fragestellungen in Bezug auf Big Data und Crowd Data
- Austausch zwischen Erfahrungsträger*innen aufbauen und etablieren, weitere Veranstaltungen im Sinne des Fachgesprächs durchführen
- Verstärktes Einbringen von stadtentwicklungsplanerischen Überlegungen in der Smart City-Diskussion
- Zusammenarbeit mit dem Amt für Statistik intensivieren; Diskussion von Nutzungsmöglichkeiten von Big Data in der räumlichen Stadtanalyse

8.2.3 Externe Netzwerke etablieren

Verbindung von Fachperspektive und IT-Welt

Die Stadtentwicklungsplanung kann ihr Wissen im Umgang mit Big Data stärken. Die Aufbereitung von Rohdaten und die Auswertung und Analyse sehr großer und komplexer Datenmengen dürfte derzeit noch ihre Möglichkeiten und Ressourcen übersteigen. Die Zweckmäßigkeit und Effizienz einer zumindest teilweisen verwaltungsinternen Leistungser-

bringung durch Spezialist*innen ist zu prüfen.. Das entsprechende Know-how gilt es darüber hinaus auch über externe Kontakte und Netzwerke mit Spezialist*innen aus der IT-Brancheeinzubeziehen. Aus der Verbindung von Fachexpertise und IT-Kompetenzen ergeben sich aufschlussreiche Erkenntnisse. Für den Dialog mit den IT-Fachpersonen ist es notwendig, dass seitens der Stadtentwicklungsplanung zumindest ein grundlegendes technisches Verständnis vorhanden ist. Nur so ist es überhaupt möglich, sich auf derselben Ebene miteinander auszutauschen und gemeinsam Fragestellungen zu erörtern.

Folgende konkrete Schritte sind möglich:

- Kontakte und Austausch mit Lehrstühlen an Universitäten und Hochschulen etablieren, die sich spezifisch mit Themen von Big Data auseinandersetzen (Datenanalyse und-visualisierung, Digitalisierung)
- Master- und Semesterarbeiten anregen und für eigene Erkenntnisbildung und methodisches Verständnis nutzen
- Kontakte zur Startup-Szene in Berlin herstellen / forcieren
- Dialog mit der Community aufbauen (z. B. Datenjournalist*innen)
- Veranstaltungen organisieren, co-organisieren oder sponsern und Leitfragen eingeben, Ideenwettbewerbe für spezifische Fragen ausschreiben
- Bezüglich sensorgenerierten / transaktionsgenerierten Daten: In Austausch treten mit städtischen Betrieben, Betrieben in öffentlicher Hand (Strom, ÖPNV etc.)
- Beteiligung an nationalen / internationalen Forschungs- / Praxisprojekten

8.2.4 Konkrete Nutzungsmöglichkeiten prüfen

Konkrete Prüfung im StEP Zentren

Handfeste Erfahrungen in der Nutzung von Big Data lassen sich nur in konkreten Anwendungen sammeln. Mit Blick auf die Betrachtung des potenziellen Nutzens von Big Data für die einzelnen StEPs in Kapitel 7 bietet sich der StEP Zentren für mögliche pilothafte Anwendungen an. Aus verschiedenen, sich ergänzenden Datenquellen können Aussagen hinsichtlich der Abgrenzung, der Ausstattung und Aufenthaltsqualität in den Zentren gewonnen werden. Es kann vertieft geprüft werden, wie Daten von Fotoplattformen (Flickr), Ratingplattformen (Tripadvisor, Google Places) oder aus Mobilitätsplattformen (Uber) konkret genutzt werden können. Dabei ist insbesondere die Komplementarität mit weiteren Daten zu prüfen. Big Data kann beispielsweise zur Kalibrierung von Erkenntnissen genutzt werden.

Daneben sind weitere konkrete Schritte möglich:

- Prüfung inwieweit Big Data Quellen eine Relevanz für weitere Fachplanungen besitzen (analog zu den StEPs)
- Kontakt mit Plattformen suchen und Datenanalysen für die Stadt diskutieren (z. B. Immobilienplattformen)
- Kostengünstige Daten beschaffen und Erfahrungen in der Weiterverarbeitung und eventuell Kombination dieser Daten sammeln (ein Beispiel mit niedriger Eintrittshürde wäre Strava Heatmap)
- Dialog mit privaten Unternehmen suchen, um die Nutzbarkeit von (kommunalen) Unternehmensdaten im Rahmen von einzelnen Planungen zu prüfen

8.2.5 Erfahrungen mit Crowdsourcing sammeln

Crowdsourcing als Partizipationskanal nutzen

Neben der Nutzung von Big Data bieten Crowd Data mittels Crowdsourcing eigene und anders gelagerte Anwendungsmöglichkeiten. Aus Sicht der Stadtentwicklung ist Crowdsourcing in mehrfacher Hinsicht lohnend. Einerseits können Daten zielgerichtet erhoben und genutzt werden. Andererseits kann , in unterschiedlichen Formen ein partizipationsorientierter Dialog mit den Bürger*innen stattfinden. Je nach Anlage des Crowdsourcings kann die Stadt zudem bereits bei der Erhebung der Crowd Data stärker als bei der Nutzung von extern erhobenen Big Data ihre Interessen einbringen und die an die Bürger*innen herangetragenen Fragestellungen steuern.

Crowdsourcing als Partizipationskanal, dessen Möglichkeiten und Grenzen sich erst ansatzweise erkennen lassen, gilt es verstärkt zu nutzen und konkrete Erfahrungen zu sammeln. Das Anliegenmanagement, wie es in Berlin auch zur Anwendung kommt, ist dabei nur *eine* Möglichkeit.

Folgende konkrete Schritte sind möglich:

- Auswertung der Erfahrungen mit „Ordnungsamt Online“, Ableiten von Erkenntnissen für weitere Anwendungen
- Auswertung der Erfahrungen mit dem Citizen Science-Projekt „Portal Beee“ des Berlin-Brandenburgischen Instituts für Biodiversitätsforschung (BBIB), Ableiten von Erkenntnissen für weitere Citizen Science-Anwendungen
- Veranstaltungen organisieren oder unterstützen (etwa Civic Hacking, Civic Hackathons, wie beispielsweise „hackyourcity“)
- Initiieren eines Pilotprojektes zu Participatory Sensing (Datensammlung mittels mobiler Endgeräte)
- Testen von neuen Formen der ePartizipation im Rahmen von Planungsprozessen

- Verstärkte Bereitstellung und auch verwaltungsinterne Nutzung von Open Data, Kontakte in die Open Data-Szene nutzen, um den Datenkreislauf zu schließen[22]

[22] Vgl. zum Beispiel http://geo.ebp.ch/2016/11/15/win-win-schliessen-des-open-data-feedback-loops

A1 Literatur

Attard, M., Haklay, M., Capineri, C. (2016): The Potential of Volunteered Geographic Information (VGI) in Future Transport Systems. In: Urban Planning, Volume 1, Issue 4, 6-19.

Al Nuami, E., Al Neyadi, H., Mohamed, N, Al-Jaroodi, J. (2015): Applications of big data to smart cities. In: Journal of Internet Services and Applications.

Berner Fachhochschule (2015): Chancen, Risiken und Handlungsbedarf des Bundes.

Batty, Michael (2013): Big data, smart cities and city planning. In: Dialogues in Human Geography 3(3), 274-279.

BBSR, Bundesinstitut für Bau-, Stadt- und Raumforschung (2015): Smart Cities International – Strategien, Strukturen und Pilotvorhaben, Eva Schweitzer (ed.). Bonn: BBSR, Bundesinstitut für Bau-, Stadt- und Raumforschung.

Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e. V. (BITKOM): (2014): Big-Data-Technologien – Wissen für Entscheider. Leitfaden. Berlin.

Clark, D. (2012): Wisdom (http://www.nwlink.com/~donclark/knowledge/wisdom.html)

Culbricht, C. (2009): Screen Scraping – Wann ist das Auslesen und die Veröffentlichung fremder Daten zulässig, http://www.rechtzweinull.de/archives/100-screen-scraping-wann-ist-das-auslesen-und-die-veroef fentlichung-fremder-daten-zulaessig.html

Deutsch, A. (2014): Reiseportale dürfen Ryanair-Flugdaten auslesen, http://www.lto.de/recht/hintergruende/h/bgh-urteil-izr22412-screen-scraping-flugdaten-automatisiert-auslesen-ryanair-reiseportal

Goldsmith, S., Crawford, S. (2014): The Responsive City: Engaging Communities Through Data-Smart Governance.

Haan, Y., Höffken, S. (2015): Civic Hacking – Bürgerengagement und urbane Mitgestaltung am Beispiel "Hack your city". In: Planerin, 03/2015.

Hatch, D. (2017): Uber shares mobility data with urban planners, http://citiscope.org/story/2017/uber-shares-mobility-data-urban-planners

Höffken, S. (2014): Urbane Mitgestaltung und neue Technologien. Smarte Partizipation. In: BMBU, Das Magazin zu den Pilotprojekten der Nationalen Stadtentwicklungspolitik. Ausgabe 9, September 2014

Hollenstein, L., Purves R. (2010): Exploring place through user-generated content: Using Flickr tags to describe city cores. JOSIS, 1, 21-48.

Howe, J. (2006): The Rise of Crowdsourcing. In: wired.com.

IBM (2016): Big data & analytics hub. http://www.ibmbigdatahub.com/

Kanton Zürich (2016): Airbnb blüht vor allen in den Städten. Medienmitteilung vom 14.12. 2016. Zürich.

Kitchin, R. (2015): Big data and official statistics: Opportunities, challenges and risks. In: The Programmable City Working Paper 9.

Kitchin, R. (2014): Big Data, new epistemologies and paradigm shifts. In: Big Data & Society.

Koellreuter Ch. (2016): Das Stellengebot ein Jahr nach der Aufhebung des Mindestkurses des Schweizer Frankens. Basel.

Louail, T., Lenormand, M., Cantu Ros, O.G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J., Barthelemy, M. (2014): From mobile phone data to the spatial structure of cities. In: Physics & Society, Scientific Reports 4.

Magistrat der Stadt Wien (2014): Smart City Wien Rahmenstrategien. Wien: Magistrat der Stadt Wien.

Missing Maps (2017): Missing Maps, http://www.missingmaps.org

OpenStreetMap Wiki (2017): Missing Maps Project, http://wiki.openstreetmap.org/wiki/Missing_Maps_Project

Purves, R., Edwardes, A.J., Woods, J. (2011): Describing place through user-generated content, First Monday, 16(9), http://pear.accc.uic.edu/ojs/index.php/fm/article/view/3710/3035

Ruths, D., Pfeffer, J. (2014): Social media for large studies of behavior. In: Science, Vol. 346, Issue 6213, 1063-1064.

SATW Schweizerische Akademie der Technischen Wissenschaften (2017): Ethische Herausforderungen für Unternehmen im Umgang mit Big Data. Zürich.

Solmecke, C. (2013): Ist Screen Scraping legal?, https://www.wbs-law.de/urheberrecht/ist-screen-scraping-legal-39848

Stadt Zürich, Stadtentwicklung (2013): ePartiziaption in der Stadtentwicklung. Begriffe – Möglichkeiten – Empfehlungen. Zürich.

Straumann, R. (2012): Chancen und Risiken von Crowdsourcing für die öffentliche Verwaltung, http://geo.ebp.ch/2012/01/24/chancen-und-risiken-von-crowdsourcing-fur-die-offentliche-verwaltung

Straumann, R.K., Çöltekin, A., Andrienko, G. (2014): Towards (re)constructing narratives from georeferenced photographs through visual analytics. The Cartographic Journal, 51(2): 152–165.

UNECE (2013): Classification of Big Data (http://www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data)

UNECE (2016): Outcomes of the UNECE Project on Using Big Data for Official Statistics.

Ushahidi (2017): Ushahidi - Read the crowd, https://www.ushahidi.com

Ward J., Barker A. (2013): Undefined By Data: A Survey of Big Data Definitions. School of Computer Science. University of St Andrews, UK.

Yuan, Y., Raubal, M. (2012): Extracting clustered urban mobility and activities from georeferenced mobile phone datasets. UCLA, Santa Barbara.

A2 Übersicht Anwendungsbeispiele

	Ziel	verwendete Daten	weiterführende Links
Fotoplattformen			
Locals and Tourists by Eric Fisher	Visualisierung der Orte an denen Einheimischen und Touristen ein Foto gemacht haben	Flickr, Picasa	https://www.flickr.com/photos/walkingsf/sets/72157624209158632/
Geotaggers' World Atlas	Visualisierung der Bewegungen im Raum während der Aufnahme von Flickr-Fotos	Flickr	https://www.mapbox.com/blog/geotaggers-world-atlas/ http://www.citylab.com/tech/2015/04/mapping-every-citys-most-scenic-routes-one-photo-at-a-time/391775/
Exploring places through user-generated content	Beschreibung und Abgrenzung von Stadträumen anhand der "Tags" hochgeladener Flickr-Bilder	Flickr, Geo-graph	http://www.josis.org/index.php/josis/article/viewArticle/13
Phototrails	Visualisierung und Analyse von Fotos nach bestimmten Kategorien und Filtern	Instagram	http://phototrails.net/instagram-cities/ http://firstmonday.org/ojs/index.php/fm/article/view/4711/3698
Breitbandaus-bau Berlin	Visualisierung und Beschreibung des Breitbandausbaus sowie die Nutzung von Instagram und Twitter in Berlin	Instagram, Twitter	http://www.breitband-berlin.de/
Ratingplattformen			
Big Data in Big Apple	Analyse und Visualisierung von Check-Ins bei einer Veranstaltung in New York	Foursquare	http://blog.vint.sogeti.com/?p=5345 http://engineering.foursquare.com/2012/09/19/big-data-and-the-big-apple-understanding-new-york-city-using-millions-of-check-ins/
Social Media			
Twitter-Maps by Eric Fisher	Visualsisierung der Orte an denen Einheimische und Touristen einen Tweet gemacht haben	Twitter	https://www.mapbox.com/labs/twitter-gnip/locals/#5/38.000/-95.001 http://firstmonday.org/ojs/index.php/fm/article/view/3710/3035 https://www.theguardian.com/news/datablog/2012/nov/22/using-twitter-mapping-urban-planning-transport-networks
One million tweet map	Interaktive Karte mit Ort und Inhalt von Tweets	Twitter	http://onemilliontweetmap.com/ http://www.citylab.com/housing/2015/03/why-most-twitter-maps-cant-be-trusted/388586/
We Feel	Interaktive Wiedergabe des emotio-nalen Zustandes der Welt und von Regionen	Twitter	http://wefeel.csiro.au/#/ http://ergonomics.uq.edu.au/iea/proceedings/Index_files/papers/1025.pdf
Sentiment NYC	Wiedergabe der Emotionen in NYC	Twitter	http://necsi.edu/research/social/newyork/newyorksentiment.pdf

GPS-Tracking			
Strava/ Run-keeper Heat Map	Visualisierung der zurückgelegten Wege von Sportlern	Strava, Run-keeper	http://labs.strava.com/heatmap/#12/8.10733/51.49547/orange/bike http://www.citylab.com/commute/2014/12/mapping-the-paths-most-traveled/383940/ http://www.dailymail.co.uk/sciencetech/article-2553352/Where-does-city-run-Beautiful-new-maps-reveal-popular-routes-world.html https://www.wired.com/2014/06/strava-sells-cycling-data/
Navigationssysteme			
Waze Connec-ted Citizens	Echthtzeit-Informationen über Vor-fälle auf der Straße	Daten werden von Usern einer Community übermittelt z.B. via Live Feed API übermittelt	https://www.waze.com/de http://www.newyorker.com/business/currency/waze-and-the-traffic-panopticon https://www.fastcompany.com/3045080/waze-is-driving-into-city-hall https://www.ft.com/content/c3a7f4c4-bae8-11e5-b151-8e15c9a029fb#axzz40sDc6rhY
Suchportale			
GoogleTrends	Informationen zu Suchbegriffen von Nutzern	Google	https://www.google.de/trends/
Mobilitätsplattformen			
City Flow Map	Visualisierung und Analyse der zu-rückgelegten Strecken	Uber	https://newsroom.uber.com/uberdata-mapping-a-citys-flow-using-ubers-ridership-data/ http://labs.sogeti.com/mapping-a-citys-flow-using-uber-data/
Unterkünfte			
WG-Gesucht Crawl	Abbildung der aktuellen Wohnge-meinschafts-Situation in Berlin	WG-Gesucht	http://ini20.de/2014/09/wg-gesucht-crawl/
Airbnb	Visualisierung von Airbnb Daten	Airbnb	http://www.airbnbvsberlin.de/ http://karat.studio/blog/zweckentfremdung

A3 Beispiele für Crowdsourcing

Weitere Beispiele aus Berlin

Leises Berlin

Im Rahmen der Öffentlichkeitsbeteiligung zum Lärmaktionsplan Berlins hatten alle interessierten Bürger*innen im Frühjahr 2013 unter der Internetadresse www.leises.berlin.de die Möglichkeit sich zu informieren und von Verkehrslärm besonders belastete Orte in Berlin zu benennen. Anhand der Fragen „Wo ist es laut?" und „Wie wird es leiser?" konnten lärmbelastete Orte lokalisiert und Vorschläge gemacht werden, wie der Lärm reduziert werden könnte. Die öffentlich sichtbaren Lärmhinweise der Bürger*innen konnten dabei zudem kommentiert und bewertet werden. Nach der aktiven Dialogphase wurden alle Lärmhinweise und die dazugehörigen Kommentare von der Senatsverwaltung Berlin zusammengefasst, ausgewertet und nach Bezirken sortiert aufbereitet. Somit hat diese Form der Beteiligung dazu beigetragen, dass die Öffentlichkeit auf schnelle und detaillierte Art und Weise Informationen bereitstellen konnte. Mittlerweile ist der Lärmaktionsplan in Kraft getreten. Der Prozess sowie alle eingegangen Lärmhinweise sind auch weiterhin auf der Internetseite abrufbar.

Abbildung 21: Vorschlag im Rahmen der Öffentlichkeitsbeteiligung (Quelle: leises.berlin.de)

Fahrradkarte per Crowdsourcing

Vom 15. bis 29. Juli 2012 machte das BMW Guggenheim Lab (eine mobile Forschungseinrichtung) Halt in Berlin. Im Zentrum der Workshops und Vorträge stand dabei das Thema der aktiven Partizipation und wie die Bürger*innen selbst bei der Entwicklung der Stadt mitgestalten können. In diesem Zuge wurde von der australischen Verkehrsplanerin Rachel Smith, zusammen mit John Schimmel von der New York University und dem Grafikdesigner Dave Dawson eine der ersten Fahrradwege-Karten

weltweilt auf Basis von Crowdsourcing entwickelt. Bei diesem Projekt konnten Radfahrer*innen ohne großen Aufwand Straßen in Berlin auf einer Google Map markieren und im Hinblick auf die Radfahrfreundlichkeit und Sicherheit bewerten. So konnten die User beispielsweise angeben, ob sie sich beim Radfahren auf der von ihnen verzeichneten Straße "zufrieden" oder "gestresst" fühlten. Nach Beantwortung mehrerer Fragen wurde die Straße dann auf einer großen Karte entweder grün oder rot markiert, wobei die grün die fahrradfreundlichen Wege darstellt. So entstand nach und nach eine von Nutzer*innen entwickelte Karte von geeigneten Fahrradwegen in Berlin. Allerdings wurde die Website eingestellt.

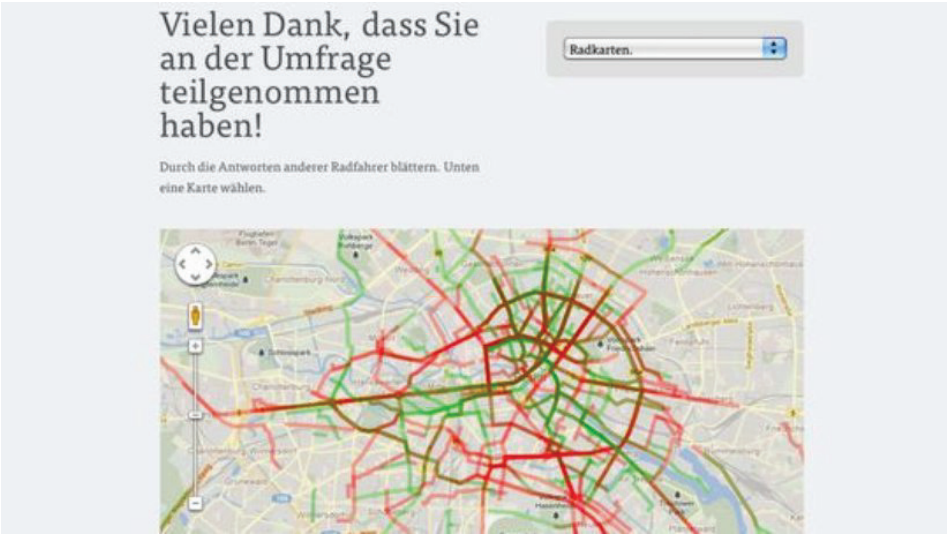


Abbildung 22: Interaktive Fahrradkarte (Quelle: http://www.zeit.de/auto/2012-07/fahrradkarte-berlin-crowdsourcing)

Portal Beee

Das Citizen Science-Projekt „Portal Beee“ wurde vom Berlin-Brandenburgischen Institut für Biodiversitätsforschung (BBIB) initiiert mit dem Ziel eine Wissensbasis rundum das Thema Biodiversität zu schaffen. Dementsprechend können Berliner*innen und Brandenburger*innen Beobachtungen übermitteln und kleinere Experimente durchführen. Beispielsweise können bei dem Projekt „Igel in der Stadt“ beobachtete Igel gemeldet werden, in dem der genaue Ort auf einer Karte lokalisiert wird und Merkmale wie Datum, Uhrzeit, Anzahl der Tiere sowie sonstige Bemerkungen (z. B. Verletzungen, Jungtiere, Auffälligkeiten im Verhalten) erfasst werden. Auf diese Weise können große Datensätze zusammengetragen werden. Das Portal bietet darüber hinaus viele Informationen rund um die laufenden Projekte und Berichte über Schlussfolgerungen aus der Daten-Auswertung. Aktuell gibt es außer dem Projekt „Igel in der Stadt“, weitere Projekte zu den Themen Wildschweine, Füchse, Gewässer und Verlust der Nacht.

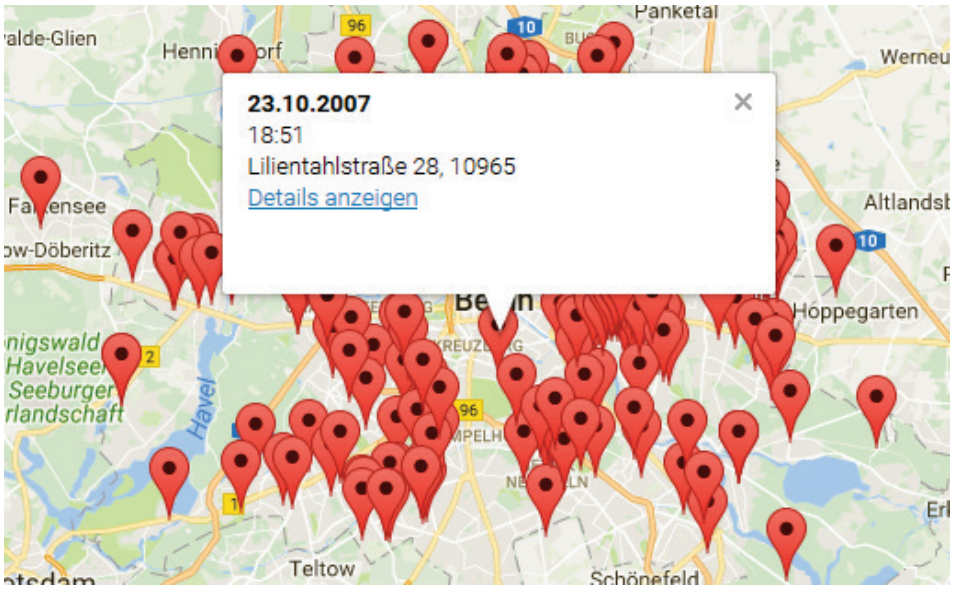


Abbildung 23: beobachtete Igel in Berlin (Quelle: http://www.portal-beee.de/igel-beobachtungen.html)

Weitere Beispiele

Geruchsmanagement Kelsterbach bei Frankfurt a.M.

In der Kleinstadt Kelsterbach bei Frankfurt am Main gab es Geruchsbelästigungen durch die Stadtentwässerung. Um die Beschwerdelage und die Geruchsbelästigungen managen zu können, wurden die Bewohner in den Jahren 2014 bis 2015 aufgerufen, ihre Geruchsempfindungen auf einer Webplattform mit Lage und Uhrzeit anzugeben. Damit wurde die Ermittlung möglicher Geruchsquellen unterstützt. Gleichzeitig konnte die Situation vor und nach Inbetriebnahme von technischen Maßnahmen gemessen und damit die Wirksamkeit der Maßnahmen festgestellt werden. Die Webplattform ist heute nicht mehr aktiv, jedoch kann die Projektdokumentation auf der Internetseite abgerufen werden.

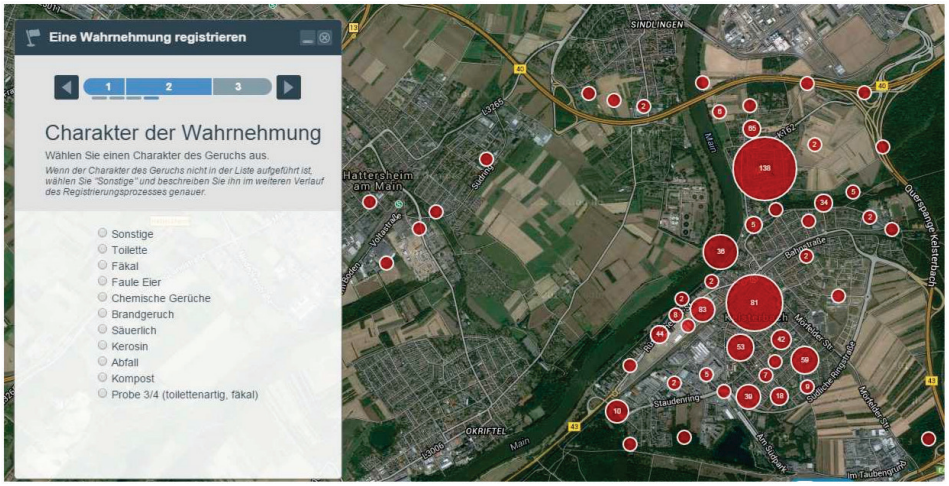


Abbildung 24: „smell map" Kelsterbach (Quelle: http://stadtentwaesserung-frankfurt.de/images/stories/dateien/P14-050_051_SEF_Kelsterbach_Rev03Kurz_barrierefrei-signed.pdf)

„Züri wie neu“

In Zürich kann man über die Plattform „Züri wie Neu“ Schäden an der städtischen Infrastruktur melden. Dies kann auf einer Website oder noch bedienungsfreundlicher direkt über eine App gemacht werden. Dabei wird im ersten Schritt der aktuelle Ort über die Ortungsdienste oder per Adresseingaben erkannt. Als nächstes kann ein Foto des Schadens aufgenommen und hochgeladen werden. Im letzten Schritt kann der Schaden zusätzlich kategorisiert und beschrieben werden (Abfall / Sammelstelle, Beleuchtung / Uhren, Brunnen / Hydranten, Graffiti, Grünflächen / Spielplätze, Signalisation / Lichtsignal, Straße / Trottoir / Platz, VBZ / ÖV). Das System hat Erfolg – bisher wurden etwa 8.900 Meldungen registriert. Die Erfahrungen zeigen auch, dass die Schäden in der Regel rasch behoben werden.

Weitere Beispiele

Frankfurt am Main: Einbringen von Projektideen und Anliegen in der Stadt Frankfurt am Main
http://www.frankfurt-gestalten.de/

Atlanta: Ermitteln von Defekten an städtischer Infrastruktur
http://www.peds.org

Boston: Mehrere Anwendungen zu verschiedenen stadtspezifischen Problemen in Boston
https://www.boston.gov/departments/innovation-and-technology/apps

Haiti: „Krisen-Mapping“:
https://www.ushahidi.com/blog/2010/04/14/crisis-mapping-haiti-some-final-reflections

London/Barcelona: Erfassung von Gerüchen („smell maps“)
http://www.independent.co.uk/news/uk/home-news/researchers-compile-smell-maps-of-urban-areas-using-the-noses-of-thousands-of-volunteers-10314480.html#gallery

Zürich: Ermittlung von Infos zu Toiletten in Zürcher Restaurants und Bars
https://www.stadt-zuerich.ch/portal/de/index/ogd/anwendungen/2014/crowdpee.html

Chicago: Ermitteln der Nachfrage nach neuen Bikeshare-Standorten
http://suggest.divvybikes.com

San Francisco: Ermitteln von Mobilitätsmustern für Planungszwecke
http://www.sfcta.org/modeling-and-travel-forecasting/cycletracks-iphone-and-android

Global: Ermitteln von Infos für eine barrierefreie Mobilität
https://news.wheelmap.org/

A4 Beispiele für Small Data

Beispiele aus Berlin

Im Kontext der Recherche wurden auch Beispiele von Datenquellen mit Bezug zur Stadtplanung Berlin identifiziert, die kaum Charakteristika von Big Data aufweisen. Folgerichtig lassen sie sich eher unter „Small Data" einordnen. Die folgenden Small Data-Quellen wurden identifiziert:

- Blogs mit Bezug zu Berlin: z. B. „Berlin, Ick liebe dir"[23], „Stil in Berlin"[24], „Muxmäuschenwild"[25]
- Websites mit zielgruppenspezifischen Informationen: z. B. „Senioren in Berlin"[26]
- Datenjournalismus, welcher die Möglichkeiten neu verfügbarer Daten nutzt und diese in innovativen Visualisierungsformen aufbereitet, z. B. von der Berliner Morgenpost[27]
- Nachbarschaftsapps zur Förderung lokaler Netzwerke durch gemeinsamen Austausch und Unternehmungen: z. B. „Nachbarschaft.net"[28], „WirNachbarn"[29], „Roundhere"[30], „Auf halber Treppe"[31], „Polly & Bob"[32]

Komplementierung oder Kontrastierung von Erkenntnissen aus Big Data

Für spezifische Fragestellungen können diese Small Data-Quellen interessante Informationen liefern, insbesondere auch zu bestimmten Gruppen der Stadtgesellschaft. Sie können beispielsweise Big Data-Untersuchungen oder Stichprobenerhebungen von Behörden ergänzen.

[23] https://berlin-ick-liebe-dir.de
[24] http://www.stilinberlin.de
[25] http://www.muxmaeuschenwild-magazin.de
[26] http://home.meinestadt.de/berlin/senioren
[27] http://www.morgenpost.de/interaktiv
[28] https://nachbarschaft.net
[29] https://wirnachbarn.com
[30] http://www.roundhere.me
[31] http://aufhalbertreppe.de
[32] https://www.pollyandbob.com

A5 Sammlung Datenquellen

1	**Nutzergenerierte Daten**	
1.1	Fotoplattformen	
	Flickr	flickr.com
	Geograph DE	http://geoen.hlipp.de
	Pinterest	https://de.pinterest.com/
	Instagram	https://www.instagram.com/?hl=de
	Amazon Photos	https://www.amazon.com/photos/home
	Photobucket	http://photobucket.com/
	Shutterfly	https://www.shutterfly.com/
	googlephotos	https://www.google.com/photos/
1.2	Ratingplattformen	
	Tripadvisor	https://www.tripadvisor.de/
	Foursquare	https://de.foursquare.com/
	Yelp	https://www.yelp.de
	Google Places	https://www.google.com/business/
1.3	Kartenportale	
	OpenStreetMap	https://www.openstreetmap.org/
	Berlinmap.org	http://berlinmap.org/de/index.html
	Mapillary	https://www.mapillary.com/
	Google Streetview	https://maps.google.com
	visitberlin	http://www.visitberlin.de/de
1.4	Social Media	
	Twitter	https://about.twitter.com/company
	Facebook	www.facebook.com
	Instagram	https://www.instagram.com/?hl=de
	Whatsapp	https://www.whatsapp.com/?l=de
	Google+	https://plus.google.com/
	Snapchat	https://www.snapchat.com/l/de-de/
	Jodel	https://jodel-app.com/
1.5	GPS Tracking	
	Strava	https://www.strava.com/?hl=de
	Runkeeper	https://runkeeper.com/
	RideWithGPS	https://ridewithgps.com/
1.6	Suchportale	
	Google	www.google.de
	Yahoo	https://de.yahoo.com/
	Bing	https://www.bing.com/
1.7	Wikis	
	Wikipedia	https://www.wikipedia.de
1.8	Datingportale	
	Tinder	http://tinderpc.com/
	Lovoo	https://de.lovoo.com/
	Love scout 24	https://www.lovescout24.de
	Parship	https://singles.parship.de/
	Elite Partner	http://www.elitepartner.de
	Grindr	http://www.grindr.com
1.9	Businessnetzwerke	
	Xing	https://www.xing.com/
	LinkedIn	https://de.linkedin.com/

2	**Transaktionsgenerierte Daten**	
2.1	Mobilitätsplattformen	
	Uber	https://www.uber.com/de-CH/
	Nextbike	http://www.nextbike.de/de/berlin/
	Sharoo	http://www.sharoo.com/
	Multicitiy carsharing	https://www.multicity-carsharing.de/
	cambio carsharing	http://www.cambio-carsharing.de/berlin
	Blablacar	https://www.blablacar.de/
	Besser Mitfahren	https://www.bessermitfahren.de/
	Drive2Day	http://www.drive2day.de/
	Berlin Shuttle	http://www.berlinshuttle.de/home.html
	Stadtmobil	http://berlin.stadtmobil.de/
	Greenwheels	https://www.greenwheels.com/de/
	Flinkster	https://www.flinkster.de/
	MyMobilityMap	https://www.mymobilitymap.de/carsharing-berlin
	Moovel	https://www.moovel.com/de/de
	DriveNow	https://de.drive-now.com/
	Car2Go	https://www.car2go.com/DE/de/berlin/
	Ubeeqo	https://www.ubeeqo.com/de/carsharing
	Emio	https://www.emio-sharing.de/
	Coup	https://joincoup.com/
	Tamyca	https://www.tamyca.de/
	Drivy	https://www.drivy.de/
2.2	Immobilienportale	
	Immobilienscout 24	https://www.immobilienscout24.de/
	immonet	www.immonet.de
	Immowelt	www.immowelt.de
	WG-Gesucht	http://www.wg-gesucht.de/
	wunschgrundstück	www.wunschgrundstück.de
	Ohne-Makler	https://www.ohne-makler.net/
	my next home	http://my-next-home.de/
	Immoexperten	https://www.immoexperten.de/
2.3	Unterkünfte	
	Airbnb	www.airbnb.de
	Booking	http://www.booking.com/index.de
	ebookers	https://www.ebookers.de/
	Kayak	https://www.kayak.de/
	wimdu	http://www.wimdu.de/
	visitberlin	http://www.visitberlin.de/de
2.4	Ausschreibungsplattformen	
	BUND Verwaltung online	http://www.bund.de/Content/DE/Ausschreibungen
	Auschreibungen.de	http://www.ausschreiben.de/
	Vergabe 24	https://www.vergabe24.de/
2.5	Tauschplattformen	
	Craigslist	https://berlin.craigslist.de/
	tauschring	http://www.tauschring.de/hauptseite.php
	Tausch dich Fit	https://www.tausch-dich-fit.de/
2,6	Jobbörsen	
	stellenangebote.de	http://www.stellenangebote.de/
	stellenanzeigen.de	http://www.stellenanzeigen.de/
	jobs.de	http://www.jobs.de/
	Monster	http://www.monster.de/
	Stepstone	https://www.stepstone.de/

2.7	Online-Handel	
	Amazon	https://www.amazon.de/
	Ebay	http://www.ebay.de/
2.8	Unternehmensdaten	
	…	
2.9	Mobilitätsdaten	
	Deutsche Bahn	https://www.bahn.de/
	Berliner Verkehrsbetriebe	http://www.bvg.de/de/
	Verkehrsbetriebe Berlin-Brandenburg	http://www.vbb.de/de
	Taxi Fare Finder	https://www.taxifarefinder.com/
3	**Sensorgenerierte Daten**	
3.1	Navigationssysteme	
	TomTom	https://www.tomtom.com/de_ch/
	Garmin	http://www.garmin.com/de-DE
	Waze	https://www.waze.com/de/
3.2	Mobilfunkdaten	
3.3	Überwachungskamerasysteme (CCTV)	
	…	
3.4	Gebäudedaten	
3.5	Biometrische Daten	

A6 Beschreibung ausgewählter Bezugskanäle

Ausgewählte Bezugskanäle werden gemäß folgender Punkte beschrieben:

Zugänglichkeit (Accessibility)

- *Bezugskanäle*: Dabei wird bei nutzergenerierten sowie transaktionsgenerierten Daten das Vorhandensein einer sogenannten Programmierschnittstelle (bzw. Application Programming Interface, API), der Möglichkeit des Datenkaufs bei der Datenquelle und bei Datenbrokern oder Angeboten der Community geprüft.
- *Kosten*, z. B. Lizenzkosten
- *Datenschutz*: Nutzungseinschränkungen

Qualität (Accuracy)

- *Objektivität*: Können die zu generierenden Informationen als repräsentativ angesehen werden oder besteht die Gefahr von Verzerrungen (z. B. ungleichmäßige Nutzung durch verschiedene Bevölkerungsgruppen)? Besteht die Gefahr von Manipulation durch die Bereitsteller der Informationen?
- *Nachhaltigkeit / Längsschnittfähigkeit*: Seit wann gibt es den Anbieter? Gibt es Aussagen zur Beständigkeit des Angebots (bspw. Service Level Agreement, SLA)? Bestehen bereits Zeitreihen und wenn ja, wie lange? Wie konstant werden Daten eingespeist (Aktualität)? Ist mit einer Fortführung der Seite / Quelle zu rechnen?
- *Qualitätssicherung*: Nennt der Datenlieferant explizit Schritte zur Qualitätssicherung (Bsp.: Minderung des Contributor Bias, Filterung von zu hohen Geschwindigkeiten bei Fußgängertrajektorien)? Welche?

Auflösung (Precision)

- *Räumliche Auflösung*: Aussagen zur räumlichen Verortung sowie zu allfälligen Unsicherheiten im Raumbezug (punktgenau, Hektar, Quadratkilometer, Bezirk, PLZ, Bei Rasterdaten: nominelle Rasterauflösung)
- *Zeitliche Auflösung*: Was ist die feinste Zeiteinheit, die die Daten unterscheiden (bspw.: Hotelgäste pro Nacht, Touristen pro Hektar pro 5-Minutenschritt, Flickr-Fotos pro Jahr)? Können Daten zeitspezifisch abgerufen werden (z. B. alle Taxifahrten bei Nacht, am Tag des EM-Finals etc.)
- *Inhaltliche Auflösung*: Inhaltlicher Detaillierungsgrad: Welche Messniveaus und bei kategorialen Daten wie viele Ausprägungen werden generiert bzw. können abgeleitet werden (Bsp.: Personen pro Hektar vs. Personen 0-u20 Jahre / 20-u40 J. / 40-100 J. pro Hektar)

Datenquelle			Datenanbieter		Daten	Zugänglichkeit		
Nr.	Datenquelle	Kat.	Datenanbieter	Kat. Daten-anbieter	Daten	Bezugskanäle	Kosten	Datenschutz / Nutzungseinschränkungen
1	Airbnb	2.3	www.airdna.co	Datenbroker	Marktreports (Revenue per available room (RevPAR), Average daily rate (ADR), Belegung, monatliche Angebots-/Nachfragetrends), Preis-vorhersage/Erlösmanagement (Angebot-/Nachfrage- und Preisprognose über 6 Monate), Market Intelligence-Report (Belegung, jährlicher Erlös, Saisonalität, Top Performers; über 12 Monate historisch), Property-Daten im Excel-Format (40 Metriken pro Airbnb-Liegenschaft)	Download von aggregierten Daten, "Rohdaten" (z.T. im Excel-Format) und Reports. Bereitstel-lung eines interaktiven airdna-Cockpits bzw. -Dashboards.	müssen angefragt werden	keine nennenswerten Einschränkungen in https://www.airdna.co/terms und in https://www.airdna.co/privacy-policy.
2	Flickr	1.1	https://www.flickr.com/services/developer	Plattform	Angaben zu Fotos (Tags, Beschreibungen, EXIF-Daten, bei Geocodierung: Ort, etc.), Usern, Useraktivitäten (Kommentare, Fotouploads, Favorites), Fotokollektionen, Gruppen.	Flickr API	keine	v.a. applikationsbezogene Einschränkungen, vgl. https://www.flickr.com/services/api/tos
3	Twitter	1.4	https://dev.twitter.com	Plattform	Angaben zu Tweets (Zeitpunkt, User, URLs, Hashtags, Text, bei Geocodierung: Ort, etc.), Twitter-Usern und deren Following-Beziehungen (Netzwerk) und Aktivitäten (Likes, Retweets).	Twitter APIs: REST API, Streaming API	keine	vgl. Policy (https://dev.twitter.com/overview/terms/policy.html); u.a. Verbot der Nutzung von "standalone location data" (https://dev.twitter.com/overview/terms/policy.html#b-maintain-the-integrity-of-twitters-products), Dritten dürfen Tweets und User nur via ID weitergegeben werden (nicht mit dem gesamten Inhalt). Die API erlaubt nur eine gewisse Abfragefrequenz, die für abfrageintensive Analysen (z.B. Ermittlung von Netzwerkbeziehungen) zum Teil sehr eng ist.
4	Instagram	1.1	https://www.insta-gram.com/developer	Plattform	Angaben zu Fotos und Videos (Tags, Beschrei-bungen, bei Geocodierung: Ort, etc.), Usern, Usernetzwerken und -aktivitäten (Kommentare, Likes).	Instagram API	keine	vgl. Platform Policy (https://www.instagram.com/about/legal/terms/api/)

Datenquelle		Qualität			Auflösung		
Nr.	Datenquelle	Objektivität	Nachhaltigkeit, Längsschnittfähigkeit	Qualitätssicherung	Räumliche Auflösung	Zeitliche Auflösung	Inhaltliche Auflösung
1	Airbnb	Vermutlich hoch (Airdna verfügt über sämtliche Airbnb-Daten). Contributor Bias unbekannt, aber auch nicht relevant für Verwendung der Daten.	Die Daten gehen insgesamt zurück bis August 2014. Airdna bemüht sich, die räumliche Abdeckung auf weitere Weltgegenden und Städte auszudehnen. Die politische Entwicklung rund um Airbnb in Europa – und somit auch jene von Airdna – ist nicht klar.	Diverse Qualitätssicherungsmaßnahmen (z.B. Filtern inaktiver Objekte) sind unter https://www.airdna.co/methodology erwähnt. Geblockte vs. gebuchte Tage wurden mit einem Machine Learning-Ansatz abgeschätzt (erläutert unter https://www.airdna.co/methodology).	Daten in ausgesuchten Weltgegenden und Städten (je nach Produkt), aggregiert auf Ebene Stadt, PLZ, Nachbarschaft (unklar, welches Level in Europa), Lage von Airbnb-Objekten vermutlich auf einige Dutzend Meter genau.	Die Daten sind tageweise für jedes Objekt erhältlich. Standardmässig werden die Entwicklungen über das letzte Jahr ausgewertet und Prognosen über 6 Monate erstellt.	Die Objekte (Unterkünfte) werden nach Anzahl Zimmer, Kosten für die Unterkunft und Bewertungskategorie unterschieden.
2	Flickr	starker Contributor Bias (vergleichsweise wenige Personen generieren einen substanziellen Teil des Inhalts)	Wachsende Anzahl neuer Fotos (durch wachsende Nutzendenbasis und/oder durch zunehmen Fotohäufigkeit), zunehmende Zahl von Mobiltelefon-Fotos (dadurch potenziell mehr geocodierte Fotos).	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit der Geocodierung von Bildern abhängig von der Geocodierungsquelle.	Daten sind zeitlich beliebig genau aufgelöst.	Rohdaten (Medien, Beschreibungen, Tags)
3	Twitter	Twitter wird deutlich weniger breit genutzt als beispielsweise Facebook. Insbesondere Personen des öffentlichen Lebens (Berühmte Personen, Politikerinnen/Politiker, Journalistinnen/Journalisten) dürften auf der Plattform übervertreten sein. Twitter verfügt über einen grossen Contributor Bias (vergleichsweise wenige Personen generieren einen substanziellen Teil des Inhalts).	Die Funktionalität von Twitter hat sich über die Zeit nur wenig verändert. Twitter ist ökonomisch noch zu wenig erfolgreich. Immer wieder gibt es Übernahmegerüchte, deren Implikationen für die Plattform unklar sind.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit der Geocodierung von Tweets abhängig von der Geocodierungsquelle. Nur wenige Prozent (i.d.R. < 3%) der Tweets sind geocodiert.	Daten in Echtzeit verfügbar (per Abfrage via REST API oder per Streaming via Streaming API).	Rohdaten (Texte, Medien, Beziehungen, URLs, Hashtags)
4	Instagram	Instagram ist die am häufigsten genutzte Fotoplattform weltweit mit hoher Aktivität. Trotz ihrem starken Contributor Bias (vergleichsweise wenige Personen generieren einen substanziellen Teil des Inhalts).	Die Anzahl Fotos pro Zeiteinheit wächst; durch wachsende Nutzerbasis und/oder zunehmende Fotohäufigkeit.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit der Geocodierung von Bildern abhängig von der Geocodierungsquelle.	Daten in Echtzeit verfügbar.	Rohdaten (Medien, Kommentare, Likes, Tags, Beziehungen)

Datenquelle			Datenanbieter		Daten	Zugänglichkeit		
Nr.	Datenquelle	Kat.	Datenanbieter	Kat. Datenanbieter	Daten	Bezugskanäle	Kosten	Datenschutz / Nutzungseinschränkungen
5	Strava	1.5	http://metro.strava.com	Plattform/ Datenbroker	Angaben zu Bewegungsmustern (Orte, Strecken) durch GPS-Tracking. Kategorisierung nach Art der Aktivität (z.B. Laufen, Radfahren), Querschnittszahlen auf Strecken, Frequenzen und Wartezeiten an Kreuzungen, Start- und Zielorte.	Download von aggregierten Daten, Daten für Weiterverwendung in Geoinformationssystemen aufbereitet	Lizenzgebühr abhängig von der Nutzerzahl der Strava App in der Region	keine nennenswerten Einschränkungen; vgl. https://www.strava.com/privacy und https://www.strava.com/terms
6	Uber	2.1	https://movement.uber.com/cities	Plattform/ Datenbroker	Verkehrsdaten aller getätigten Uber-Fahrten (Fahrtzeit, Strecke, Start- und Zielort).	Für Zugang ist bisher noch Anmeldung mit E-Mail notwendig. Uber will die Daten von Uber Movement aber bald allgemein freigeben.	keine	Verkehrsdaten werden anonymisiert und außerdem unter einer sogenannten "Creative-Commons-Lizenz" freigegeben, im Zuge derer die kommerzielle Nutzung verboten ist.
7	Foursquare	1.2	https://developer.foursquare.com	Plattform	Angaben zu Check-Ins von Nutzern (mittels Swarm), beliebte Orte, Bewertungen, Netzwerke	Foursquare APIs: Foursquare Venues API, Foursquare Standard API	keine	v.a. applikationsbezogene Einschränkungen; vgl. https://developer.foursquare.com/overview/community
8	Gnip	1.4	https://gnip.com http://support.gnip.com/gnip2.0	Datenbroker	Ermöglicht die Aggregation, Filterung und Aufbereitung von Daten aus diversen sozialen Netzwerken	Erschließung kann über Complete Access Sources oder Public API Sources erfolgen. Complete Access Sources ermöglicht dabei einen vollständigen Zugriff auf den kompletten Datenstrom, der alle öffentlichen Informationen integriert. Public APIs werden von den jeweiligen sozialen Plattformen wie Facebook angeboten und sind durch Volumen- und Häufigkeitseinschränkungen in ihrer Themenabdeckung limitiert.	je nach Bedarf unterschiedlich, Abonnement beginnt bei 500 $	vgl. Policy Twitter https://twitter.com/privacy?lang=de

Datenquelle		Qualität			Auflösung		
Nr.	Datenquelle	Objektivität	Nachhaltigkeit, Längsschnittfähigkeit	Qualitätssicherung	Räumliche Auflösung	Zeitliche Auflösung	Inhaltliche Auflösung
5	Strava	Strava Metro wird von Strava selbst betrieben und verfügt über sämtliche Daten. Die Nutzerbasis von Strava ist aber natürlich nur ein Teil der Personen, die den genannten Aktivitäten nachgehen. Wie repräsentativ diese Stichprobe ist, ist nicht bekannt.	Daten gehen zurück bis 2014. Die Anzahl der Nutzer ist steigend, somit werden künftig auch mehr Daten hochgeladen und generiert.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit abhängig vom GPS-Gerät.	Daten können bis auf Stundenebene aufgelöst werden. Standardmässig werden die Entwicklungen über das letzte Jahr ausgewertet.	Daten ermöglichen eine Tiefenanalyse von Radfahrer- und Fußgängeraktivitäten, einschließlich populärer oder vermiedener Routen, Pausen- und Wartezeiten sowie Ursprungs- / Zielzonen.
6	Uber	Uber Movement wird von Uber selbst betrieben und verfügt über sämtliche Daten. Die Nutzerbasis von Uber ist aber natürlich nur ein Teil von Personen, die ähnliche Dienste benutzen. Wie repräsentativ diese Stichprobe ist, ist nicht bekannt.	Noch keine Aussagen möglich. Das Angebot wurde erst im Januar 2017 lanciert und dient vermutlich nicht kommerziellen Zwecken sondern der Verbesserung der Reputation des Unternehmens.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit abhängig vom GPS-Gerät.	Daten bis auf Tagesebene verfügbar.	Verkehrsaufkommen zu bestimmten Tages- und Jahreszeiten exakt zu bestimmen, die durchschnittliche Fahrzeit von einem zum anderen Punkt einzusehen oder neuralgische Punkte im Stadtverkehr zu identifizieren.
7	Foursquare		Andere Apps laufen Foursquare den Rang ab. Die Entfernung des Check-In Features und die Aufspaltung in zwei mobile Apps hat bei Nutzern überwiegend negative Reaktionen hervorgebracht.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich in Abhängigkeit vom GPS-Gerät. Durch Check-In wird direkter Bezug zu einem Ort hergestellt.	Daten in Echtzeit verfügbar.	Rohdaten (Texte, Bewertungen, Medien)
8	Gnip	Objektivität je nach Plattform differenziert.	Hohe Nachhaltigkeit bzw. Längsschnittfähigkeit, da die Daten stetig archiviert werden.	Die Formate der über die Complete Access Scources empfangenen Daten sind zum Teil unterschiedlich und werden durch den Gnip Data Collector in eine einheitliche Struktur überführt.	Potenziell Meterbereich. Genauigkeit abhängig vom GPS-Gerät.	Echtzeit- und historische Daten verfügbar	Mehrere Produkte verfügbar. Firehose ermöglicht den vollständigen Zugriff auf alle Echtzeitdatenströme und die Anreicherung mit zusätzlichen Kontextinformationen. Bei PowerTrack können die Datenbestände nach Kriterien wie z.B. Geoinformationen,Vorhandensein von speziellen Stichwörtern etc. gefiltert werden. DataCollector stellt einen einheitliche API für den Zugriff auf bis zu sechs unterschiedliche APIs her.

Datenquelle			Datenanbieter		Daten	Zugänglichkeit		
Nr.	Datenquelle	Kat.	Datenanbieter	Kat. Datenanbieter	Daten	Bezugskanäle	Kosten	Datenschutz / Nutzungseinschränkungen
9	Datasift	1.4	http://datasift.com http://dev.datasift.com	Datenbroker	Ermöglicht die Aggregation, Filterung und Aufbereitung von Daten aus diversen sozialen Netzwerken	Daten können via API bezogen werden. Anwender erhält die Möglichkeit, Streams mit den für ihn relevanten Datenquellen und Filterbedingungen zu konfigurieren und anschließend auszuführen.	kostenfreier Testzugang mit einigen Funktionseinschränkungen möglich; je nach Bedarf sind weitere Kosten unterschiedlich; Abonnement beginnt bei 3.000 $	vgl. Policy http://datasift.com/terms/#privacy-policy
10	google-places	1.3	https://developers.google.com/places	Plattform	Detaillierte Orts- und Sachinformationen, insbesondere zu Unternehmen und Sehenswürdigkeiten, aus Google Maps- und der Google+-Datenbank. Die Angaben könnten unter anderem Kontaktinformationen, Reviews, Fotos, Preisniveau und Öffnungszeiten umfassen. Mit Stand Februar 2017 sind die "beliebten Besuchszeiten" (popular times), die in Google Maps angezeigt werden via API *nicht* verfügbar. Diese Funktion ist aber ein Feature, das von vielen Entwickler*innen gewünscht ist. Dieser Wunsch wurde von Google aufgenommen, aber noch nicht umgesetzt.	GooglePlaces API	Die Grenze der kostenfreien Nutzung liegt bei 150.000 Anfragen pro 24 Stunden. Bei höherem Bedarf kann man eine Lizenz erwerben.	vgl. https://developers.google.com/places/web-service/usage?hl=de

Datenquelle		Qualität			Auflösung		
Nr.	Datenquelle	Objektivität	Nachhaltigkeit, Längsschnittfähigkeit	Qualitätssicherung	Räumliche Auflösung	Zeitliche Auflösung	Inhaltliche Auflösung
9	Datasift	Objektivität je nach Plattform differenziert.	Hohe Nachhaltigkeit bzw. Längsschnittfähigkeit, da die Daten stetig archiviert werden. Die Reichweite der zu erschließende Daten ist allerdings je nach Datenquelle variabel.	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Potenziell Meterbereich. Genauigkeit abhängig vom GPS-Gerät.	Echtzeit- und historische Daten verfügbar	Verfügt über ein Modul zur Aggregation von Daten mit einem umfangreichen Filter-System, wodurch Daten nach Bedeutung und Kontext kategorisiert werden können.
10	google-places	Der Prozess der Aufnahme von Lokalitäten in Google Places ist nicht ganz klar. Besitzer von Geschäften können unter anderem die Aufnahme via Google My Business beantragen. Die Abdeckungsgüte ist unbekannt und sicher räumlich variabel.	Vermutlich hoch	Es sind keine Qualitätssicherungsmaßnahmen bekannt.	Meterbereich	Aktualität abhängig von den anbietergeprüften Einträgen und den nutzermoderierten Beiträgen.	Rohdaten (Koordinaten, Fotos, Beschreibungen etc.)

A7 Fachgespräch

Am 17. Februar 2017 fand in Berlin ein Fachgespräch mit Teilnahme von Vertreter*innen der Auftraggeberin, der Auftragnehmerin und einem erweiterten Kreis von Sachverständigen statt.

Tabelle 8: Teilnehmende des Fachgesprächs

Teilnehmer*in	Institution
Dietrich Bangert	Senatsverwaltung für Stadtentwicklung und Wohnen
Stephan Egerer	Senatsverwaltung für Stadtentwicklung und Wohnen
Paul Hebes	Senatsverwaltung für Stadtentwicklung und Wohnen
Magdalena Konieczek-Woger	Senatsverwaltung für Stadtentwicklung und Wohnen
Kurt Nelius	Senatsverwaltung für Stadtentwicklung und Wohnen
Elke Plate	Senatsverwaltung für Stadtentwicklung und Wohnen
Thorsten Tonndorf	Senatsverwaltung für Stadtentwicklung und Wohnen
Franziska Fronzke	Senatsverwaltung für Umwelt, Verkehr und Klimaschutz
Hartmut Kästner	Senatsverwaltung für Umwelt, Verkehr und Klimaschutz
Ulrich Binner	Bezirksamt Tempelhof-Schöneberg
Sandy Arndt	Bezirksamt Treptow-Köpenick
Hartmut Bömermann	Amt für Statistik Berlin-Brandenburg
Benjamin Seibel	Technologiestiftung Berlin
Jens Libbe	Deutsches Institut für Urbanistik
Christof Abegg	EBP Schweiz AG
Holger Frantz	EBP Deutschland GmbH
Ralph Straumann	EBP Schweiz AG
Alain Thierstein	EBP Schweiz AG
Tobias Voigt	EBP Deutschland GmbH