

Open Data und die Risikowahrnehmung in der Öffentlichen Daseinsvorsorge

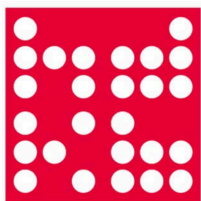
Prof. Rebecca D. Frank, PhD

RA Prof. Dr. Max von Grafenstein LL.M

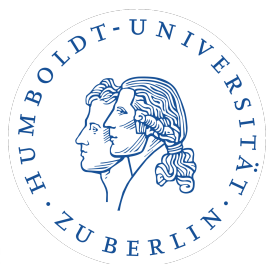
Laura Rothfritz, M.A.

24. Januar 2022

**EINSTEIN
CENTER**
Digital Future



Universität der Künste Berlin



Ein Forschungsprojekt des Einstein Center Digital Future durchgeführt von Rebecca D. Frank (Institut für Bibliotheks- und Informationswissenschaft an der Humboldt-Universität zu Berlin) und Max von Grafenstein (Universität der Künste Berlin).

Inhaltsverzeichnis

Executive Summary	1
1 Einführung und Hintergrund	2
2 Methodik und Vorgehen	3
3 Ergebnisse	5
3.1 Datenkataloge und Datenklassifikationen	5
3.2 Open Data	6
3.3 Risikoeinschätzungen für Daten aus den Unternehmen	10
4 Zusammenfassung: Herausforderungen und Hürden	12

Executive Summary: Empfehlungen

Trotz gesetzlicher Verpflichtungen und Erwartungen der Öffentlichkeit Daten offen zu stellen, sind viele Unternehmen im Bereich der öffentlichen Daseinsvorsorge bisher zurückhaltend, ihre Daten zu veröffentlichen. Die vorliegende Studie möchte den Gründen einer solchen Zurückhaltung am Beispiel von Infrastrukturversorgern in einer deutschen Großstadt nachgehen. Die Erkenntnisse sollen eine Basis dafür liefern, geeignetere Maßnahmen zur Veröffentlichung oder zum Teilen von Daten zu entwickeln, von denen nicht nur die Öffentlichkeit, sondern auch die öffentlichen Stellen bzw. Unternehmen selbst profitieren können. Dabei dient diese Studie als ein wichtiger Schritt, um bestehende Herangehensweisen und Herausforderungen im Bezug auf diese Maßnahmen zu identifizieren und einzuordnen. Aufbauend auf den Ergebnissen der Studie ergeben sich folgende Empfehlungen für Infrastrukturversorger, die Unternehmensdatensätze als Open Data veröffentlichen möchten oder dies bereits tun:

- Aufgrund der bestehenden heterogenen Ausprägung eines grundlegenden Verständnisses des Konzeptes "Open Data" innerhalb der am Datenmanagement beteiligten Akteure wird dringend empfohlen, **mit Hilfe von Schulungen ein einheitliches Verständnis von Anforderungen, Definitionen und Zwecken von Open Data zu schaffen**. Es sollte von höchster Priorität sein, dass Open Data nicht bedeutet: ein internes Teilen von Daten, die Preisgabe von Daten an einzelne Unternehmen oder die Veröffentlichung von Datentabellen in Berichten. Daten liegen als Open Data vor, wenn sie der Öffentlichkeit diskriminierungsfrei unter Wahrung des gesetzlichen Rahmens zur freien Nutzung, Weiterverbreitung und -verwendung zur Verfügung gestellt werden. Hierzu zählen auch die Verwendung von offenen Lizenzen und maschinenlesbaren, nicht-proprietären Dateiformaten. Nur wenn alle verantwortlichen Stakeholder und Akteure in den Unternehmen auf dem gleichen Wissensstand sind, können weitere organisatorische und strategische Planungen zu Open Data erfolgen.
- In so gut wie allen Unternehmen gibt es wenig bis gar keine übergreifende Sichtweise darauf, welche Daten an welchen Stellen in welcher Form vorliegen. Zwar gibt es innerhalb der Fachbereiche oder Untereinheiten in den Unternehmen häufig bereits ausgereifte technologische und organisatorische Prozesse im Datenmanagement, jedoch fehlen übergreifende Strukturen. Es wird daher dringend empfohlen, in allen Unternehmen sowohl die **Entwicklung von unternehmensweiten Datenkatalogen** voran zu treiben oder zu starten als auch eine **strukturierte und formalisierte Datenklassifikation vorzunehmen**. Nur auf Grundlage einer genauen Übersicht, welche Daten überhaupt in welcher Form vorhanden sind und für welche Zwecke unter welchen Umständen genutzt werden dürfen, können Entscheidungsprozesse für eine eventuelle Veröffentlichung von geeigneten Datensätzen als Open Data angestoßen werden.
- Zuletzt müssen **konsistente, standardisierte Prozesse und Richtlinien für Entscheidungen über die Freigabe von Daten** geschaffen werden. Hierzu gehören nicht nur technologische Workflows sondern auch eine **klare Aufgaben- und Verantwortungszuweisung für designierte Personen** und/oder Rollen in den einzelnen Unternehmen. Hierzu bedarf es in vielen Unternehmen sicherlich einer Art Kulturwandel, denn dezidierte Open Data Strategien bringen organisatorischen und personellen Aufwand mit sich. **Dies muss unbedingt durch Entscheidungsträger in der Unternehmensführung unterstützt werden.**

Einführung und Hintergrund

In ihrer *EU data strategy* berechnet die EU-Kommission den Wert der Datenökonomie im europäischen Binnenmarkt im Jahr 2025 auf ca. 829 Billionen € und damit auf ca. 5,8 % des in der EU erzielten Bruttonutzenprodukts. Mit den geeigneten politischen Maßnahmen und Investitionen sollen ca. 4-6 Milliarden € in gemeinsame europäische Datenräume und einen europäischen Zusammenschluss von Cloud-Infrastrukturen und -Diensten investiert werden.¹

Eine dieser politischen Maßnahmen ist die gesetzliche Verpflichtung öffentlicher Stellen und bestimmter öffentlicher Unternehmen, grundsätzlich alle in ihrem Besitz befindlichen Dokumente öffentlich zu stellen. Ausgenommen sind insbesondere solche Dokumente, deren Veröffentlichung gegen Datenschutz- oder IT-Sicherheitsgesetze, Gesetze des geistigen Eigentums oder für den Schutz von Geschäftsgeheimnissen verstoßen würde.² Auch auf nationaler Ebene gibt es nicht nur ähnliche Datenstrategien der Bundesregierung, sondern eine zunehmende Zahl an Informationsfreiheitsgesetzen, die mit einem vergleichbaren Schema die Veröffentlichung oder zumindest das Teilen von Daten vorsehen.³

Ungeachtet weiterer bereits in Planung befindlicher Gesetzgebungsinitiativen, die den Anwendungsbereich solcher Datenteilungspflichten zunehmend ausdehnen, ist auch der öffentliche Druck gerade auf öffentliche Stellen und öffentliche Unternehmen groß, ihre Daten freiwillig zu veröffentlichen. Begründet wird diese Forderung insbesondere damit, dass diese Stellen bzw. Unternehmen durch öffentliche Gelder finanziert werden und daher ihre Daten im Gegenzug der Öffentlichkeit zur Verfügung stellen sollen.⁴ Trotz gesetzlicher Verpflichtungen bzw. einer öffentlichen positiven Wahrnehmung beobachtet man in vielen öffentlichen Stellen sowie Unternehmen eine große Zurückhaltung, Daten zu veröffentlichen.

Die vorliegende Studie möchte den Gründen einer solchen Zurückhaltung am Beispiel von Infrastrukturversorgern in einer deutschen Großstadt nachgehen. Die Erkenntnisse sollen eine Basis dafür liefern, geeignetere Maßnahmen zur Veröffentlichung oder zum Teilen von Daten zu entwickeln, von denen nicht nur die Öffentlichkeit, sondern auch die öffentlichen Stellen bzw. Unternehmen selbst profitieren können. Dabei dient diese Studie als ein erster wichtiger Schritt, um bestehende Herangehensweisen und Herausforderungen im Bezug auf diese Maßnahmen zu identifizieren und einzuordnen.

1. https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283.

2. Art. 1 der RICHTLINIE (EU) 2019/1024 DES EUROPAISCHEN PARLAMENTS UND DES RATES vom 20. Juni 2019 über offene Daten und die Weiterverwendung von Informationen des öffentlichen Sektors (Neufassung).

3. Siehe das vom 13. Dezember 2006 (BGBl. I S. 2913), geänd. durch G v. 8.7.2015 (BGBl. I S. 1162) oder auch das Gesetz zur Förderung der Informationsfreiheit im Land Berlin (Berliner Informationsfreiheitsgesetz - IFG) vom 15. Oktober 1999, zuletzt neu gefasst durch Artikel 5 des Gesetzes vom 12.10.2020 (GVBl. S. 807).

4. https://de.wikipedia.org/wiki/Open_Data#Forderungen_der_Open-Data-Bewegung.

Methodik und Vorgehen

Ziel dieser Studie war es, herauszufinden, wie in öffentlichen Versorgungsinfrastrukturen mit Daten umgegangen wird, welche Erfahrungen und Herausforderungen im Hinblick auf eine Öffnung von Daten bestehen könnten und inwiefern für das Datenmanagement und die Datensicherheit verantwortliche Personen in den Unternehmen Chancen und Risiken zu Open Data absehen können. Hieraus ergaben sich folgende übergeordnete Forschungsfragen, die die Studie inhaltlich prägten:

Forschungsfragen

Wie verstehen die beteiligten Akteure in den Unternehmen "Open Data" und wie stehen sie der Öffnung von Daten ihres eigenen Unternehmens gegenüber?

Welche Risiken oder Hürden im Bezug auf die Öffnung von Daten werden von den Verantwortlichen in den Unternehmen gesehen?

Kongruent mit dem qualitativen Forschungsansatz dieser Studie beruhen die Ergebnisse auf leitfragengestützten, semi-strukturierten Interviews. Hierfür wurde zuerst ein Interviewleitfaden erstellt, der Fragen zu folgenden Themenkomplexen enthielt:

- Fragen zu bestehenden Aspekten und/oder geplanten Entwicklungen im Bereich Datenmanagement, Data Sharing und Open Data in den einzelnen Unternehmen
- Fragen zur Einschätzung des Wertes (finanziell, ideell, anderweitig) der Daten in den Unternehmen
- Fragen zur Beurteilung von Risikofaktoren und Herausforderungen im Hinblick auf die Bereitstellung von Open Data durch die Unternehmen

Es wurden insgesamt 18 Interviews zwischen September und November 2020 geführt, die zwischen 30 Minuten und einer Stunde dauerten. Interviewpartner wurden mit Hilfe eines "purposive convenience sampling" Ansatzes ausgewählt. Diese Art der Auswahl beschreibt einerseits die begründete Form der Auswahl von Personen anhand ihrer Rollen (purposeful) und andererseits die Auswahl nach Zugangsmöglichkeit (convenience) sowie durch das Schneeballprinzip.

Insgesamt nahmen **sieben Personen aus einem Unternehmen für die Grundwasserversorgung, fünf Personen aus einem Unternehmen für die Stadtreinigung, und sechs Personen aus Unternehmen für die Stromversorgung** an der Studie teil. Alle Personen sind in ihren Unternehmen mit dem Umgang mit Daten vertraut und/oder beauftragt, die jeweiligen Rollen der Interviewten wurde von uns gemäß Tabelle 1 kategorisiert.

Alle Interviews wurden nach Zustimmung der Teilnehmenden als Audiodatei aufgezeichnet und anschließend transkribiert. Die Transkripte wurden mit Hilfe der Software für qualitative Datenanalyse MAXQDA kodiert und nach Methoden der qualitativen Inhaltsanalyse ausgewertet. Für die Analyse wurden die verwendeten Codes sowohl induktiv als auch deduktiv basierend auf Konzepten aus der Literatur, aufbauend auf thematischen Schwerpunkten des

Tabelle 1: Rolle der Teilnehmenden in den Unternehmen

Rolle der Teilnehmenden	Anzahl der Interviews
IT/Infrastruktur	4
Kundendaten	4
Asset Data	4
Recht/Data Governance	5
Innovation	1
Gesamt	18

Interviewleitfadens sowie wiederholt auftauchenden Aspekten aus den Interviews erstellt und angewendet. Unter anderem wurden folgende Codes innerhalb folgender Kategorien vergeben: Open Data, Data Sharing, Data Reuse, Data Governance und legale Aspekte (inkl. DGSVO), Dateninfrastrukturen, Data Value, Risiken und Vorteile von Open Data.

Ergebnisse

Die Ergebnisse dieser Studie lassen sich unter folgenden Hauptaspekten zusammenfassen:

1. Existierende und/oder geplante Infrastrukturen und Systeme für das Datenmanagement, vor allem in Hinblick auf Datenkataloge und Datenklassifikation,
2. Verständnis und Bewertung von Open Data,
3. Einschätzungen der Teilnehmenden zu Risiken für die Daten ihrer Unternehmen sowie im Bereich Open Data.

Datenkataloge und Datenklassifikationen

In fast allen Unternehmen besteht die Problematik, dass eine übergeordnete Sicht oder Auskunftsmöglichkeit, welche Daten in welcher Form and welcher Stelle vorliegen, z.B. als Datenkatalog, fehlt. Hieraus resultieren sogenannte Datensilos, also abgeschirmte Datenbestände in den jeweiligen Abteilungen der Unternehmen, die mit Datenbeständen aus anderen Abteilungen häufig nicht interoperabel oder vernetzbar sind und die sich zwangsläufig nur in der Obhut der in der Abteilung jeweils für das Datenmanagement zuständigen Person (falls vorhanden) befinden.

Nur zwei Teilnehmende gaben an, dass es in ihrem Unternehmen Datenkataloge gäbe, die durch die Datenschutzbeauftragten oder durch SAP verwaltet werden. Jedoch wurde von sechs teilnehmenden Personen die bevorstehende oder angelaufene Planung und Entwicklung solcher Kataloge bejaht und stark befürwortet.

"Wir haben kein durchgängiges IT-System, im Moment. Da sind wir auf dem Weg dorthin. Um letztendlich dann auch die Daten in einem System und in einem Workflow-System zu haben."(Interviewpartner:in 08)

"Also es gibt ein Tool und es gibt auch ein Projekt, was eine Struktur vorsieht mit Data Stewards und Data Analysten und Data Owner, die sozusagen dann das alles katalogisieren. Da gibt es jetzt ein, zwei Beispiel-Bereiche, in denen das gemacht wird. Aber das ist noch ein ganz langer Weg bis man da, sage ich mal, ansatzweise ein Teil des Geschäfts oder ein Großteil des Geschäfts überhaupt in dem System hat." (Interviewpartner:in 10)

"Aber zusätzlich kommt halt dazu, dass, was wir gerade schon angesprochen haben, mit einem Daten-Katalog und einer, wie soll ich sagen, einer Landkarte, wo ich was an Daten finde, das wird halt in der Zukunft noch einen viel stärkeren Wert spielen, dass ich die ganzen Daten, die ich jetzt digitalisiert habe oder auch gerade noch digitalisiere in den letzten Jahren, dass ich die halt wiederfinde und nicht mehr sozusagen meine Sachbearbeiter habe, die irgendwo sagen "Ja, ja, das habe ich hier in meinem Ordner und guck mal, habe ich"ja, so. Diese Rolle von den Experten, die wissen, wo die Daten liegen, die wird sich ja ganz stark wandeln. Und damit verbunden ist halt dieses Auffinden von Daten und dieser, soll ich sagen, neue Wert, die Daten überhaupt zu finden, um sie nutzen zu können. Das ist halt für mich noch mal eine andere Wert-Definition." (Interviewpartner:in 10)

Auf der anderen Seite gaben alle Teilnehmenden an, dass in ihren Unternehmen Klassifizierungssysteme für Datenbestände, insbesondere für sensible Daten, existieren. Dies wurde als überaus wichtig bewertet, gerade im Bezug auf Datenbestände, die unter die DSGVO oder andere Datenschutzbestände fallen, sowie für Datenbestände, die sich auf Aspekte der kritischen Infrastruktur oder der Wettbewerbsfreiheit beziehen.

"Dass es finalisiert ist, weiß ich nicht, aber ich weiß, dass man daran arbeitet, an einer Datenklassifizierung, ob man tatsächlich dort ... also, zumindest möchte man die Daten clustern, klassifizieren und dann mit einem entsprechenden Attribut versehen, wie diese Daten ... wie man mit diesen Daten verfahren darf. Das ist auch zwingend erforderlich. Ich würde sagen, das braucht jedes Unternehmen, sonst befindet man sich irgendwie im rechtsleeren Raum und weiß letztendlich gar nicht, wie man damit umgehen darf oder sollte, hm." (Interviewpartner:in 03)

Datensilos, fehlende Datenkataloge und übergeordnete Datenmanagementzuständigkeiten führen zudem auch dazu, dass der Austausch von Daten zwischen Unternehmensabteilungen sehr häufig nicht automatisiert werden kann, sondern abhängig ist von informellen Beziehungen zwischen einzelnen Personen und auf Grundlage von Nachfragen via E-Mail oder Telefon angeregt wird.

"Aber letztendlich hier hat der Fachbereich die Informationen, welche Information ihm zur Verfügung gestellt werden müssen, damit sie dann arbeitsfähig sind. Also das ist natürlich so. Also das war dann vielleicht ein bisschen, ich habe mir sozusagen so eine große, große Übersicht vorgestellt, die Sie jetzt fragen, wo alle Daten drin stehen. Das haben wir in dieser Weise nicht. Aber die Fachbereiche wissen natürlich welche Daten sie haben." (Interviewpartner:in 09)

"Dann geht momentan viel über die IT-Architekten und über die IT-Bereiche und über System-Betreuer oder System Owner. Und so ist sozusagen der informelle Weg, weil die Daten ja in den Systemen hängen und dann fragt man einfach immer die, die die Systeme betreuen, wie die Daten aussehen, weil halt also einfach in Ermangelung, sage ich mal, immer aktuell gepflegten Landkarte." (Interviewpartner:in 10)

"Also ich kenne da natürlich sehr viele Personen und ansonsten wird eben über die jeweilige Leitung der Organisationseinheit angefragt: Habt ihr da was dazu oder wie sieht das aus? Ja. Ist also kein...bisher kein formaler Prozess, welche Daten jetzt wie klassifiziert oder gehandhabt sind." (Interviewpartner:in 05)

Generell werden Entwicklungen im Bereich Datenklassifizierung sowie Datenkatalogen von allen Befragten als überaus wichtig angesehen. Besonders Klassifizierungssysteme werden als unerlässlich betrachtet, um Entscheidungen treffen zu können, welche Arten und Kategorien von Daten überhaupt für eine Veröffentlichung als Open Data vorgesehen werden können. Bestehende Projekte haben hierbei häufig auch einen besonderen Fokus auf sensible Daten im Rahmen der DSGVO.

Open Data

Für Unternehmen, die einen Auftrag haben, Teile ihrer Daten öffentlich zur Verfügung zu stellen, ist es von größter Wichtigkeit, dass alle an diesem Prozess beteiligten Personen das

gleiche Verständnis von Open Data besitzen. In dieser Studie zeigte sich, dass dies bei den beteiligten Unternehmen nicht durchgehend der Fall ist. Im Gegenteil: Auf die Frage, was die Teilnehmenden unter Open Data verstehen und wie Daten offengelegt werden könnten wurden sehr heterogene Antworten gegeben. Obwohl viele interviewte Personen ein grundsätzliches Verständnis für Open Data mitbrachten, schwankte das Wissen zu genauen Ausprägungen und Voraussetzungen von Open Data zwischen sehr detailliertem Fachwissen und vagem Teilverständnis. So herrschte bei manchem Beteiligten keine grundlegende Unterscheidung zwischen Open Data, also Daten die für jeden auffindbar, verwendbar und nachnutzbar in maschinenlesbarer Form veröffentlicht wurden, und Daten, die (intern) mit anderen geteilt wurden oder die sich in Dokumenten wie Berichten oder anderen Veröffentlichungen befinden, vor.

“Wir öffentlichen Unternehmen dürfen jetzt nicht durch eine Auflage unsere Unternehmensdaten transparent zu machen, um sie allgemein nutzbar zu machen, in einen Wettbewerbsnachteil rutschen, der uns dadurch entsteht, dass mögliche Informationen, die wir haben, kostenlos preisgegeben werden müssen und andere darüber einen Gewinn erzielen. Da bin ich dagegen.” (Interviewpartner:in 02)

“Ich könnte mir nur was drunter vorstellen und sagen, es gibt also irgendwie einen Pool von Daten, auf denen jemand, der berechtigt ist, zugreifen kann.” (Interviewpartner:in 06)

“Ich glaube eine Ahnung zu haben, was Open Data ist. Open Data ist zum einen wortwörtlich übersetzt offen auf Daten zugreifen zu können, die Möglichkeit zu bieten, Daten aus unterschiedlichen Quellen gemeinsam zu nutzen. Um auf dem Weg für die Beteiligten einen größeren Nutzen für ihren jeweiligen Zweck zu erzeugen.” (Interviewpartner:in 08)

“Also Open Data bedeutet, dass Daten zur Verfügung gestellt werden und nutzbar gemacht werden und auffindbar gemacht werden. Das heißt, bei Open Data sind Daten in einem Katalog katalogisiert, Metadaten-Katalog katalogisiert, beschrieben, mit einer offenen Lizenz versehen und die Daten selbst werden in einem offenen Format angeboten, sodass sie dann auch genutzt werden können.” (Interviewpartner:in 10)

“Natürlich, das ist ja eigentlich Open Data, ist für mich die Frage: Wie werde ich eigentlich meiner Verantwortung der Stadt und anderen Stakeholdern gegenüber gerecht? Indem ich diese Monopolsituation, die ich irgendwo habe, indem ich die natürlich auch im Sinne der Stakeholder ein Stück weit öffne. Und da, wo ich Dinge wirklich sharen kann, es tatsächlich auch offen tue.” (Interviewpartner:in 01)

Trotz einiger Erfahrung mit Open Data in fast allen Unternehmen, erwähnten so gut wie alle interviewten Personen eine nachhaltig bestehende Problematik von unterschiedlichen Erwartungen aus der Öffentlichkeit und Politik: einerseits wird von zukunftsweisenden Unternehmen erwartet, zumindest Teile Ihrer Daten zu veröffentlichen, und andererseits müssen vertrauenswürdige Unternehmen, gerade im öffentlichen Sektor, verstärkt darauf achten, Datenschutz- und Wettbewerbsrechte sowie IT-Sicherheit zu gewährleisten.

“Ich glaube, der zweite Punkt ist der, dass, selbst, wenn man mit dem Gedanken spielt, dass man da offener mit umgeht, jetzt gerade durch das Thema Datenschutzgrundverordnung et cetera auch eine sehr große Unsicherheit entstanden ist, so dass man

im Zweifelsfalle immer lieber sagt 'Nein, also das sind jetzt Geheimnisse, das ist jetzt personalisiert' oder was weiß ich. So dass ich immer auf der... letztendlich immer auf der sicheren Seite bin." (Interviewpartner:in 14)

"Wir brauchen nicht nur eine Übersicht und wir brauchen nicht nur Klassifizierungen und sozusagen Maßnahmen. Wir brauchen auch einen irgendwie einen SINN dafür. Und ich finde, wir brauchen auch einfach mal den Blick über den Tellerrand: Wie machen es denn die anderen, ja? Was machen denn die...[...] was machen denn die anderen Kritis-Unternehmen hier in [Stadt]?" (Interviewpartner:in 18)

"Ich glaube, dass wir da in der Zeit in einer Branche sind, wo wir um genau diesen Punkt ringen. Nämlich auf der einen Seite die Erhöhung des Wertes von Daten durch ihre, ich nenne es mal, Veröffentlichung für ein Open Data. Auf der anderen Seite aber auch die Risiken, die daraus entstehen können bezogen auf den Punkt, den wir hatten mit vertraulichen Daten." (Interviewpartner:in 08)

Diese teils zutiefst widersprüchlichen Erwartungshaltungen wurde von vielen Teilnehmenden als ein großer Unsicherheitsfaktor im Bezug auf Open Data gesehen, denn Entscheidungsprozesse dazu, welche Daten auf welche Weise veröffentlicht werden können und sollten, sind häufig hoch komplex und langwierig. Daraus folgt, dass in vielen Unternehmen die Auffassung vorherrscht, dass der Aufwand, diesen Konflikt aufzulösen, in keiner Relation zum späteren Ergebnis (Open Data) stehe und das Risiko es nicht wert sei, Fehler zu begehen.

"Also zum einen natürlich, weil es sehr aufwendig ist und zum anderen, wenn man so eine kritische Infrastruktur betreibt, dann hat man immer die Möglichkeit, da vorher schon so ein bisschen die Grenzen aufzuzeigen und zu sagen, also solche Daten gebe ich definitiv nicht raus, die gehören mit zu den kritischen Daten einer kritischen Infrastruktur. Ich glaube, deshalb würde man diesen Stein auch gar nicht umdrehen unbedingt. Aber das ist meine persönliche Einschätzung." (Interviewpartner:in 03)

"Aber wir müssen immer berücksichtigen, dass das nicht unser unmittelbarer Unternehmensauftrag ist. Sofern ist Open Data für uns immer auch ein gutes Stück Stakeholder-Management und ist ein gutes Stück motiviert aus der Tatsache, dass wir natürlich im positiven Sinne eine Verantwortung für die Stadt haben." (Interviewpartner:in 01)

"Am Ende des Tages dachten wir uns so: Okay, selbst wenn wir... Was waren wir? Wir waren, glaube ich, so eine Hand voll Unternehmen. Selbst wenn WIR uns da jetzt weiter zu einsetzen, ja, irgendwie muss es so eine übergeordnete Instanz geben aus [der Politik] wahrscheinlich dann auch heraus, die sowas einfach koordiniert, die auch Vorgaben macht, weil sonst sind wir fünf von ich weiß nicht wie vielen Unternehmen, die sowas dann initiieren, aber wenn dann alle anderen irgendwie ihr eigenes Süppchen kochen oder ganz anders mit Daten auch umgehen, dann hat man nichts gewonnen." (Interviewpartner:in 07)

Dennoch erwähnten viele Teilnehmende, dass Open Data einen durchaus positiven Effekt für den Verbraucherschutz, für die Verwaltung und für (innovative) Geschäftsmodelle haben könnte. Es bestand allgemein die Ansicht, dass eine Öffnung von Daten auch den Wert der Daten, sowohl für das Unternehmen selbst als auch für die allgemeine Öffentlichkeit stark

erhöhen könnte. Personen, die in der Rolle Recht/Data Governance tätig sind (z.B. im Datenschutz) äußerten jedoch tendenziell eher negativere bzw. besorgtere Meinungen zum Thema Open Data. Durchweg wurde der Wert der Unternehmensdaten als sehr hoch angesehen und es wurde teilweise bedauert, dass Open Data Strategien in den Unternehmen nicht weitreichender unterstützt werden (z.B. durch Stellenausbau, Flexibilisierungsmaßnahmen o.ä.).

“Aber das auch nur deshalb, weil es natürlich... es gibt keine Datenkataloge bisher. Es gibt keine Datenklassifizierung, wo man auch relativ vielleicht festlegen könnte, was wäre denn geeignet, was nicht. Diesen ganzen Vorlauf, dieses ganze konzeptionelle Vordenken, das gibt es bisher noch überhaupt nicht.” (Interviewpartner:in 14)

“Weil wenn sie [X Anzahl] Mitarbeiter haben und davon [X Anzahl] klassisch ausgebildete [Beruf] sind, dann ist das nicht unbedingt ein Unternehmen, was nun eine ganz, ganz große Start-up-Mentalität hat. Und diesen Faktor darf man nicht vergessen, wenn man sich mit Themen wie Open Data, und das ist nur ein Beispiel, beschäftigt, bricht das auch ein gutes Stück diese Strukturen auf, schafft auch Möglichkeiten, innerhalb des Unternehmens dann mal bereichsübergreifend unterwegs zu sein.” (Interviewpartner:in 01)

“Ich glaube, dass die Verfügbarkeit von Daten ein wesentlicher Teil ihres Wertes darstellt. So dass ich dem folgend sagen würde, grundsätzlich erhöht sich der Wert von Daten mit ihrer Verfügbarkeit und mit ihrer Möglichkeit, sie zu nutzen. Wenn ich also Daten ohne Open Data an punktgleichen nur für MICH und nur für einen Zweck nutzen kann, dann ist das sicherlich ein großer Wert. Wenn ich aber über die Verfügbarkeitmachung erstmal im Unternehmen für andere Bereiche dort auch einen Zweck erfüllen kann, erhöht sich ihr Wert. So dass sich letztendlich der Wert mit der Anzahl ihrer Nutzung erhöht.” (Interviewpartner:in 08)

Zusammenfassend lassen sich sowohl Unsicherheiten durch die Anforderungen aus der Öffentlichkeit als auch weitere Unsicherheiten in den unternehmenseigenen Organisation bzw. Staretegieplanung, Kompetenzen und Priorisierungen identifizieren. Dies lässt sich jedoch auch daraus begründen, dass Definitionen von Open Data teilweise weit auseinander gingen und daher Einschätzungen zu Möglichkeiten, Nutzen und Hindernissen sehr unterschiedlich von den interviewten Personen eingeschätzt wurden. Unsicher bleibt auch weiterhin, inwiefern sich aus Open Data Bestrebungen ein Mehrwert für die Unternehmen generieren lassen könnte. Um eine gute Argumentationsgrundlage für die interne Durchsetzung von eventuell auftretenden Arbeitsaufwänden zu liefern, müsste dieser Mehrwert klarer definiert sein.

“Ja, da bin ich der Auffassung... diese Frage ist ein bisschen schwierig zu beantworten, weil ich verstehe diesen Open Data Ansatz schon so, dass ich zunächst mal, wenn ich weiß, dass in meiner Organisation Daten vorhanden sind, mir angucke, welche kann ich einfach mal rauslesen und zur Verfügung auch stellen, damit Dritte in die Lage versetzt werden, sich darüber einfach mal Gedanken zu machen. Das ist eigentlich mein Punkt. Also ich bin da im Moment auf so einer Schiene, wo ich sage: Ich muss ja erstmal dafür werben, dass überhaupt eine Bereitschaft dafür da ist. Und wir wissen, welche Daten könnten das denn theoretisch sein.” (Interviewpartner:in 14)

“Ich glaube, dass man die Leute aber für Innovationen erst dann begeistern kann, wenn die das Gefühl haben, dass die den Boden unter den Füßen haben bei dem,

was sie sowieso schon in ihrer Arbeit machen, ja? Also erstmal eine Sicherheit haben für das, was sie überhaupt verwalten an Daten, also generieren, verwalten. (...) Das wäre für mich der erste Schritt zu sagen: Was haben wir für Daten? Was haben wir für Klassifizierungen? Was haben wir für Bedürfnisse? Was veröffentlichen wir wo? Wie können wir bestimmte Prozesse gestalten? Und dann wäre für mich der nächste Schritt zu sagen: „Okay, und jetzt lasst uns doch noch mal nachdenken, was könnte man denn damit machen? Wo ist das Interesse der Öffentlichkeit? Wo können wir da noch Kompromisse machen und wo dann eben nach wie vor nicht?“ (Interviewpartner:in 18)

Risikoeinschätzungen für Daten aus den Unternehmen

Die Teilnehmenden identifizierten in den Interviews eine Reihe von Risiken, die sie im Bezug auf die Veröffentlichungen von Daten aus ihrem Unternehmen sahen. Diese Ergebnisse stehen jedoch auch in enger Verbindung mit dem jeweiligen Verständnis von Open Data der interviewten Personen.

Fast alle Teilnehmenden identifizierten Risiken aus folgenden Bereichen:

- Betriebs- und Geschäftsgeheimnisse: Wie kann sicher gegangen werden, dass veröffentlichte Daten nicht gegen das Betriebsgeheimnis verstoßen oder andere Unternehmen auf dem Markt (unbeabsichtigt) privilegieren und dadurch den freien Wettbewerb beeinträchtigen?
- Rechtliche Fragen und Kundendaten: Wie können Open Data und Datenschutz in Einklang gebracht werden?
- Sicherheitsfragen: Welche Daten, vor allem aus den kritischen Infrastrukturen, können überhaupt veröffentlicht werden, ohne dass die öffentliche Sicherheit oder die Infrastruktur des Unternehmens gefährdet werden?

Für Daten, die bereits als Open Data veröffentlicht wurden oder für die ein Unternehmen die Veröffentlichung als Open Data plant, wurde kein Risiko gesehen, denn die Entscheidungsprozesse zur Veröffentlichung wurden hier bereits durchlaufen. Jedoch wurde angebracht, dass ein Risiko für die Veröffentlichung von Open Data sei, dass sich der Aufwand zu mehr Transparenz nicht lohnen könnte, da es für das Unternehmen selber keine nennenswerten Vorteile bringe und dass Transparenzbestreben weiterhin nicht genug anerkannt werden.

“Und da fangen sie dann irgendwann an, zu überlegen: Ist das eigentlich klug, irgendwie alles rauszugeben? Oder machen wir es wie die anderen und schieben einfach so diese Standards nach oben?“ (...) Also da ist das, diese Haltung zur Transparenz wird sicherlich nicht belohnt zwangsläufig, ja.” (Interviewpartner:in 16)

“Ja, ist es... also ist es möglich? Na klar. Aber da muss man sich halt Gedanken zu machen und will ich es jetzt offenlegen oder will ich es nicht offenlegen. Aber solange ich es nicht offenlegen muss, muss ich mir diese Gedanken nicht machen bei jedem Punkt und das abwägen. Das kann schon aufwendig sein.” (Interviewpartner:in 16)

“Ich sehe so als... also prinzipiell, wenn ich Daten habe, die öffentlich zu machen, so wie sie sind, ist kein großer Aufwand. Ich glaube, Aufwand entsteht, wenn ich jedes

Mal prüfen muss, ergeben sich daraus Gefahren. (...) So, dann, wir brauchen eine Prüfschleife, ob daraus Risiken entstehen können, was ein Aufwandstreiber ist. Ein zweiter Aufwandstreiber ist sicherlich auch ein Stück weit, die Daten dann aufzubereiten, so dass sie für andere besser nutzbar sind. Und ein dritter Aufwandstreiber aus meiner Sicht ist dann auch, das Feedback auf diese Daten zu verarbeiten. Weil wenn Fragen eingehen, wenn sie genutzt werden, zunächst mal im positiven Sinne, das irgendwie zu verarbeiten ist sicherlich ein Aufwandstreiber. Und ehrlicherweise aber auch in die andere Richtung, wenn es dort... wenn daraus Kritik erwächst, auch mit der umgehen zu können." (Interviewpartner:in 08)

Weitere Risiken bei der Veröffentlichung von Daten der Unternehmen wurden vor allem von Teilnehmenden aus den Unternehmen der kritischen Infrastruktur (z. B. Wasser, Strom) im Bereich der Sicherheit gesehen. So sind kriminelle Aktivitäten oder Angriffe, die Unternehmensdaten verwenden, nicht auszuschließen. Desweiteren wurden große Risiken für den Datenschutz und die Sicherheit von personenbezogenen Daten (sowohl Kundendaten als auch interne Daten der Mitarbeitenden) gesehen.

Kontrollverlust ist ein weiteres Risiko, welches in den Interviews eine herausgehobene Rolle spielte. Für einige interviewte Personen negierte eine Datenveröffentlichung das Kontrollbedürfnis über unternehmenseigene Datensätze. Unkontrollierte Verwendung von diesen Daten könnte dann dem Unternehmen schaden und zudem einen großen Einfluss auf die Reputation des Unternehmens haben. Reputationsschäden durch fälschliche oder nicht durchdachte Datenveröffentlichung wurden in fast allen Interviews aufgeführt und spielt im Hinblick auf Open Data eine große Rolle.

"Ich glaube, das sind zwei Faktoren. A möchte man gerne seine Daten für sich selber behalten, das ist das eine Thema. Den Mehrwert, genau der, was könnte ein Mehrwert sein, wenn wir solche Daten veröffentlichen? Und natürlich auch die Sorge darum, dass wir hier Dinge rausgeben, die uns am Ende schaden können, dass wir es raus gegeben haben. Und das sind so die Dinge, die ich da für wichtig halte." (Interviewpartner:in 09)

"Also viel Image-Schaden, also ein großer Image-Schaden, weil es, weil man die... Je nachdem, wie man die Daten dann immer mal verknüpft, können natürlich auch falsche Aussagen dabei rauskommen. Dass Aussagen dabei rauskommen, die nicht der Realität entsprechen. Deswegen sind wir mit solchen Sachen auch immer sehr sparsam und geben eigentlich auch immer nur Daten raus, wenn wir die gleich auch kommentieren. Also unkommentierte Daten bergen eigentlich die Gefahr, dass die fehlinterpretiert werden oder in einem Kontext verfälscht werden." (Interviewpartner:in 11)

Somit besteht als eine weitere Unsicherheit, dass durch Open Data- und Transparenz-Bestreben sowohl die Reputation und das Vertrauen in ein Unternehmen gestärkt werden kann, gleichzeitig aber auch, dass durch unüberlegtes Verhalten hohe Reputationsschäden auftreten könnten.

Zusammenfassend sahen alle Teilnehmenden eine Reihe an Risiken sowohl für die Unternehmensdaten an sich als auch für eine theoretische Öffnung von Datensätzen. Viele der Risiken konnten als konkretisierte Fälle beschrieben werden. Andere bestanden aus hypothetischen Ereignissen, die durch eine unüberlegte Öffnung von Daten herbeigeführt werden könnten. Es zeigte sich in den Interviews, dass Unsicherheiten zu Möglichkeiten und Auswirkungen von Open Data für viele der Unternehmen eine hervorgehobene Rolle einnehmen, während Mehrwerte häufig weniger klar benannt werden konnten.

Zusammenfassung: Herausforderungen und Hürden

Insgesamt lassen sich dieser Studie vier übergeordnete Herausforderungen und Hürden für die Veröffentlichung von Open Data durch die untersuchten Unternehmen identifizieren:

1. Ein Mangel an Wissen und Strukturierung der verfügbaren Datensätze

In einigen Unternehmen sind Datenkataloge in Entwicklung, jedoch nicht vollständig. In fast allen Unternehmen gibt es Datensilos, in denen die zuständigen Personen isoliert agieren müssen und die eine Interoperabilität von Unternehmensdatensätzen häufig nicht gestatten.

2. Mangelndes Verständnis und/oder mangelnde Autorität im Bezug auf rechtliche Fragen im Zusammenhang mit Daten

Häufig fehlen klare Kriterien oder Verfahren für Entscheidungen über Daten und deren Veröffentlichung. Desweiteren fehlen häufig eine oder mehrere Personen mit klarer Verantwortung für Entscheidungsprozesse und Abläufe zur Datenveröffentlichung.

3. Abstraktivitätsgrad von zukünftigem Nutzen vs. konkrete Risiken

Die Vorteile des Teilens von Daten, ihres Nachnutzungspotenzials und der daraus folgenden konkreten Mehrwerte für die einzelnen Unternehmen sind häufig noch zu abstrakt, während Risiken als konkreter wahrgenommen werden. Dies führt zu einer Konzentration auf Risiken statt auf Vorteile und zu einer erhöhten Skepsis über eventuellen Nutzen sowie zu einer zurückhaltenden Haltung gegenüber Innovationspotentialen.

4 Aufwand und Arbeitsumfang für die Datenveröffentlichung

Häufig fehlt es an organisatorischen Möglichkeiten oder am Willen, den Aufwand für die Datenveröffentlichung in Unternehmen zu bewerkstelligen. Gleichzeitig gibt es große Ungewissheiten darüber, wie laufende Prozesse im Bereich Datenschutz und -sicherheit integriert werden können. Während die eigentliche Datenveröffentlichung als relativ einfach beschrieben wird, sind Entscheidungs- und Verwaltungsprozesse sehr anspruchsvoll und zeitaufwändig. Personen, die direkt im Datenmanagement involviert sind, stehen Open Data Bestrebungen häufig sehr positiv gegenüber. Es fehlt ihnen jedoch an Entscheidungsgewalt, vermeintlich "einfache" technische Lösungen umzusetzen.

Danksagung

Dieses Projekt wurde vom InfraLab Berlin und dem Einstein-Zentrum Digitale Zukunft unterstützt. Wir möchten dem Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, der Universität der Künste Berlin und allen Personen danken, die freiwillig ihre Zeit für diese Forschung zur Verfügung gestellt haben.

Projektteam

Prof. Rebecca D. Frank, PhD., Einstein Centre Digital Future Juniorprofessor für Information Management an der Institut für Bibliotheks- und Informationswissenschaft (Humboldt-Universität zu Berlin) beschäftigt sich mit der Risikokonstruktion bei der Prüfung und Zertifizierung digitaler Repositorien. Weiterhin umfasst ihre Forschung die Bereiche digitaler Langzeitarchivierung, digitaler Datenkuration und Datennutzung mit besonderem Fokus auf soziale und ethische Faktoren, die sich auf den Datenaustausch und auf die Archivierung und Wiederverwendung digitaler Informationen auswirken. Sie promovierte an der University of Michigan School of Information, wo sie zuvor auch ihr Studium mit Spezialisierung auf Digitale Langzeitarchivierung abschloss. Ihre Forschungsarbeit wurde von der National Science Foundation und der Australian Academy of Science gefördert.

Prof. Dr. Max von Grafenstein, LL.M., ist Rechtswissenschaftler, Rechtsanwalt und Serial Entrepreneur. 2018 wurde er als Professor für Digitale Selbstbestimmung an die Universität der Künste Berlin (UdK) im Rahmen des Einstein Center Digital Future (ECDF) berufen. 2016 hat er das Forschungsprogramm Governance of Data-Driven Innovation und Cybersecurity am Alexander von Humboldt Institut für Internet und Gesellschaft (HIIG) aufgebaut und leitet es bis heute mit. Sein generelles Forschungsinteresse gilt der Governance datengetriebener Innovation, Schwerpunkte liegen unter anderem auf Ko-Regulierungsstrategien sowie dem Regulierungsansatz (Data) Protection by Design.

Laura Rothfritz, M.A. ist wissenschaftliche Mitarbeiterin und Doktorandin am Lehrstuhl Information Management des Instituts für Bibliotheks- und Informationwissenschaft an der Humboldt-Universität zu Berlin. Sie erhielt ihren Bachelorabschluss im Fach Bibliotheksmanagement an der Fachhochschule Potsdam und schloss dort 2019 auch ihren Master im Fach Informationswissenschaften ab. Ihre Forschungsinteressen liegen im Bereich Forschungsdatenmanagement und Open Science mit besonderem Schwerpunkt auf sozio-technologischen Perspektiven auf kollaborative Tätigkeiten im Data Management und Data Reuse und einem starken Bezug zu STS (Science and Technology Studies). Sie promoviert zum Thema "Misstrauen in Dateninfrastrukturen" bei Prof. Rebecca D. Frank.