

Erkenntnisse aus dem Beteiligungsprozess zur Erarbeitung einer neuen Open-Data-Strategie für Berlin

29.06.2022

Das Wichtigste in Kürze

- ➔ Die automatisierte Bereitstellung von Open Data ist ein Indikator für die erfolgreiche Umsetzung von Verwaltungsdigitalisierung und des modernen Datenmanagements in der Verwaltung.
- ➔ Die Bereitstellung von Open Data braucht infrastrukturelle Grundlagen: Neben adäquater Hard- und Software ist auch eine Anpassung von Prozessen, Strukturen und Zuständigkeiten unumgänglich.
- ➔ Eine moderne Verwaltung mit komplexen Aufgaben braucht hochwertige und verknüpfte Datensätze (Linked Open Data).
- ➔ Der Berlin Data Hub ist ein wichtiger Baustein für ein modernes Datenmanagement und sollte zeitnah auf den Weg gebracht werden. Gleichzeitig muss die Strategie bereits jetzt auf Linked Open Data ausgerichtet sein und parallel mit dem geplanten Data Hub die dafür notwendige Infrastruktur aufgebaut werden.

Einleitung

Mehr als 10 Jahre nach dem ersten Open-Data-Strategiepapier überarbeitet das Land Berlin diese Strategie, um Impulse für die zukünftige Schwerpunktsetzungen zu erarbeiten. Die Federführung für den Strategieprozess liegt bei der Senatsverwaltung für Wirtschaft, Energie und Betriebe; für die Umsetzung von Beteiligungsworkshops wurde die Open Knowledge Foundation Deutschland beauftragt; die Open Data Informationsstelle Berlin (ODIS) unterstützt den Prozess.¹ Unter Beteiligung der Berliner Öffentlichkeit und des Berliner Open Data Ökosystems, bestehend aus Verwaltung, Wirtschaft, Wirtschaft und Zivilgesellschaft konnten in der ersten Jahreshälfte 2022 eine Vielzahl an Perspektiven auf den Status Quo sowie Anregungen und Impulse in Hinblick auf die zukünftige Ausrichtung von Open Data

¹ Auf der Website <https://strategie.odis-berlin.de/> wird der Strategieprozess offen und transparent begleitet. Dort ist alles Wissenswertes zur Open Data Landschaft Berlins und rund um den Prozess, die Dokumentationen und Ergebnisse der Workshops und die beteiligten Akteur:innen zu finden.

gewonnen werden. Im Vorfeld war die allgemeine Öffentlichkeit zu einer Online-Beteiligung eingeladen worden.

In diesem Dokument werden die bisherigen Kernergebnisse des Beteiligungsprozesses zusammengefasst, sowie am Ende illustrative Schlaglichter aus den Workshops skizziert, die Interesse an einer weiteren Beschäftigung, Vertiefung und Diskussion mit dem Thema wecken sollen. Workshopübergreifend haben sich drei Erkenntnisse herauskristallisiert: Open Data braucht eine neue Erzählung, die sich auf die interne Nutzung in der Verwaltung und die damit verbundene Nützlichkeit für die einzelnen Verwaltungsbeschäftigten fokussiert. Teil dieser Erzählung muss die Verknüpfbarkeit von Datensätzen über viele Verwaltungsquellen und darüber hinaus sein, um verteilte Informationen zu einem gemeinsamen, abfragbaren Wissensschatz verknüpfen zu können. Der geplante Data Hub ist wichtig, braucht klare Vorgaben und sollte möglichst modular in seinen Funktionalitäten wachsen, während gleichzeitig die Grundlagen für Linked Open Data als Default gelegt werden müssen.

Das Versprechen von Open Data muss neu justiert werden: Open Data ist Verwaltungsmodernisierung und führt zu effizienterer Arbeit

Der Mehrwert von Open Data wird vielfach mit extrinsischen Motivationsfaktoren begründet. Diese reichen von Argumentationen rund um Transparenz und Nachvollziehbarkeit von Verwaltungsentscheidungen bis hin zu ungehobenem Potenzial durch die Nutzung der Daten durch Dritte. Forschende könnten die Daten wissenschaftlich nutzen, Unternehmen könnten innovative Geschäftsideen mit den Daten entwickeln und die engagierte Zivilgesellschaft könnte Lösungsvorschläge für gesellschaftliche Probleme anhand der Offenen Daten erarbeiten. So richtig diese Argumente sind, scheinen sie dennoch in den Verwaltungen den Eindruck zu erzeugen, dass sie Open Data in erster Linie für Dritte bereitstellen und dieser erhebliche Aufwand für sie selbst keinen Nutzen bringt.

Über den gesamten Beteiligungsprozess hinaus ist deutlich geworden, dass der Verwaltung die zentrale Rolle im Open-Data-Ökosystem zukommt. So nannten bei der Online-Befragung die Mehrheit der Beteiligten mit 42 Prozent bei der Verwaltung das größte Nutzungspotenzial von Open Data. Unter den Verwaltungsakteur:innen sahen dies mit 55 Prozent sogar noch mehr. Und auch in den Beteiligungsworkshops war übergreifend Konsens, dass die Berliner Verwaltung zwar sowohl ein enormes Potenzial offener Daten besitzt als auch die größte Nutznießerin offener Daten sein kann, dieses Potenzial aber bei weitem noch nicht ausschöpft. Gleichwohl beispielsweise auf dem Berliner Open Data Portal immer mehr Datensätze aus der Berliner Verwaltung veröffentlicht werden oder Fachverfahren an das Open Data Portal angeschlossen werden, setzt sich das Prinzip *open by default*, welches auch in der Berliner Open Data Rechtsverordnung verankert ist, in den Berliner Behörden nur langsam durch. Es bedarf daher eines grundlegendes Verständnis über eine Zielrichtung und einer Herangehensweise für die Berliner Verwaltung, die über personelle, finanzielle und organisatorische Anpassungen hinaus Open Data in der Berliner Verwaltung zukünftig tragen sollen. Dabei ist klar, dass Open

Data in der Berliner Verwaltung ein Spiegelbild der generellen Data Governance der Berliner Verwaltung ist.

Die Bereitstellung von Open Data muss in der Weise in den eigenen Verwaltungsalltag integriert werden, dass sie (mittelfristig) *nicht* zu einem Doppel- oder Mehraufwand führt, sondern die eigene Arbeit effizienter und einfacher macht (insbesondere den Austausch zwischen Behörden). Um dies zu erreichen, müssen die bestehenden Fachverfahren aller Berliner Ebenen reformiert und geöffnet werden. Rahmenverträge mit Dienstleister:innen, die IT-Lösungen für Fachverfahren anbieten, müssen daraufhin überprüft werden, ob sie die Bereitstellung von Open Data in gebotener Qualität und Granularität ermöglichen (open by design) und ob sie die notwendigen Voraussetzungen für Linked Open Data schaffen.

In den Workshops wurde vielfach von Verwaltung und landeseigenen Betrieben berichtet, dass der Informationsfluss besonders dann stockt, wenn Personen auf Daten von anderen Stellen in Berlin angewiesen sind. Dort zeigt sich immer wieder, dass die angefragten Daten nicht strukturiert und somit auch nicht durchsuch- und auffindbar vorlagen, sondern beispielsweise als Aktenvermerke oder in gedruckter Form. Für die veraktende Person gibt es häufig weder Anreiz noch die notwendigen Werkzeuge, die in einem Vermerk enthaltenen Faktendaten maschinenles- und -findbar abzulegen; es genügt die Ablage in einer Papierakte. Dritte, die diese Informationen für ihre behördliche Arbeit benötigen, müssen diese Faktendaten deswegen händisch in den Akten finden und übertragen – womöglich über alle Bezirksamter hinweg.

Die Verwaltungsmodernisierung darf nicht darauf beschränkt sein, analoge Prozesse nur zu digitalisieren (z.B. von Akte zu e-Akte). Vielmehr muss es auch darum gehen, ein umfassendes Datenmanagement in der Verwaltung aufzubauen, das es ermöglicht, Prozesse neu zu definieren und Aufgaben effizienter zu erledigen. Es wird dabei immer wichtiger werden, mit strukturierten Datensätzen zu arbeiten und diese auch anderen Verwaltungen zur Verfügung zu stellen oder die Verknüpfbarkeit dieser Daten sicherzustellen.² Wenn in der Verwaltung mit Daten gearbeitet wird, müssen diese schon für eine mögliche Veröffentlichung vorbereitet werden (open by default), ohne dass dies für die Ausführenden als unverhältnismäßige Mehrarbeit empfunden wird. Wenn diese Veröffentlichung ohne weiteres Zutun der erfassenden Person geschehen kann, kann dies ein Indikator für eine gelungene IT-Strategie sein.

Wenn die Bereitstellung von Open Data durch die Berliner Verwaltung wie oben beschrieben in die eigene Verwaltungsarbeit integriert und als Teil des Berliner Datenmanagements verstanden wird, wird die Datenqualität steigen. Denn dort, wo die Verwaltung die Daten auch selbst braucht und regelmäßig mit ihnen arbeitet, erfolgt die Qualitätssicherung quasi automatisch. Dann stehen die Chancen hoch, dass auch Dritte (u.a. Wissenschaft, Wirtschaft, Zivilgesellschaft) diese bereitgestellten Daten nutzen und ihr Potenzial noch erweitern.

² Aufgrund der hohen Relevanz der Verwaltungsmodernisierung und des Entwickelns eines modernen Datenmanagements ist es wichtig, die Ansätze für Open Data auch mit den Überlegungen zu Digitalstrategie und Smart-City-Prozess abzugleichen und auszutauschen.

Mehr PDFs und bessere Excellisten reichen nicht: Eine moderne Verwaltung mit komplexen Aufgaben braucht verknüpfte Datensätze (Linked Open Data)

Open Data ist nicht gleich Open Data. Bislang messen wir den Erfolg von Open Data meist, indem wir die Anzahl von bereitgestellten Datensätzen zählen. Ob diese beispielsweise fehlerhafte PDF-Dokumente, veraltete Excellisten oder mit Metadaten versehene CSV-Dateien sind, unterscheiden wir selten. Der wirkliche Erfolg hängt aber maßgeblich auch am Reifegrad der Offenheit.

Seit vielen Jahren stecken wir bei Open Data bei einem Reifegrad zwischen "PDF" und "Excellisten" fest. Wir könnten noch viele Jahre weiter daran arbeiten, beispielsweise von "PDF" zu "barrierefreies PDF" und von "Excellisten" zu "fehlerfreie Excellisten mit Optimierungen" zu kommen. Dies wird auch einen Großteil der Arbeit in Verwaltungen ausmachen. Der Aufwand dazu ist jedoch immens und wird auch bei jedem neuen Datensatz sehr groß bleiben (bei bleibenden großen Limitationen dieser Formate in Sachen Offenheit). Die Klagen über den "großen Mehraufwand ohne spürbaren Nutzen" werden daher auch bleiben. Mit dem demografischen Wandel und der digitalen Transformation müssen die Prozesse in der Verwaltung so umgestaltet werden, dass es zu Effizienzgewinnen durch Automatisierungen kommt. Dies ist nur möglich, wenn Datensätze miteinander sprechen lernen, oder mit anderen Worten: durch Linked Open Data³. Datensätze müssen so angelegt werden, dass sie möglichst offen kompatibel, gut beschrieben und eindeutig identifizierbar sind. Nur so können Verknüpfungen automatisch generiert werden.

Ein Beispiel aus den Workshops: Während der Covid-19-Pandemie nutzten viele Datenjournalist:innen tabellarische Daten des Robert-Koch-Instituts (zum Infektionsgeschehen, zum Impfstatus sowie Geodaten aus anderen Quellen für die Zuordnung auf Bundesländer, Städte und Landkreise), um für journalistische Angebote diese Zahlen automatisiert grafisch aufzubereiten und dadurch erfahrbar zu machen (z.B. tägliches Corona-Dashboard mit Deutschlandkarte in der Süddeutschen Zeitung). Die Auswertungen konnten zwar weitgehend skriptgesteuert vorgenommen werden, den verwendeten Skripten musste jedoch händisch zugeordnet werden, in welcher Spalte der Tabelle Informationen welcher Art zu erwarten waren. Immer wenn das RKI seine Tabellen verändert hat (Spaltenüberschriften, neue Spalten, Spaltenreihenfolge), funktionierten die Skripte der Journalist:innen nicht mehr und mussten händisch erneut angepasst werden. Dies führte oft zu hektischen Telefonaten zwischen Redaktionen und RKI am frühen Morgen.

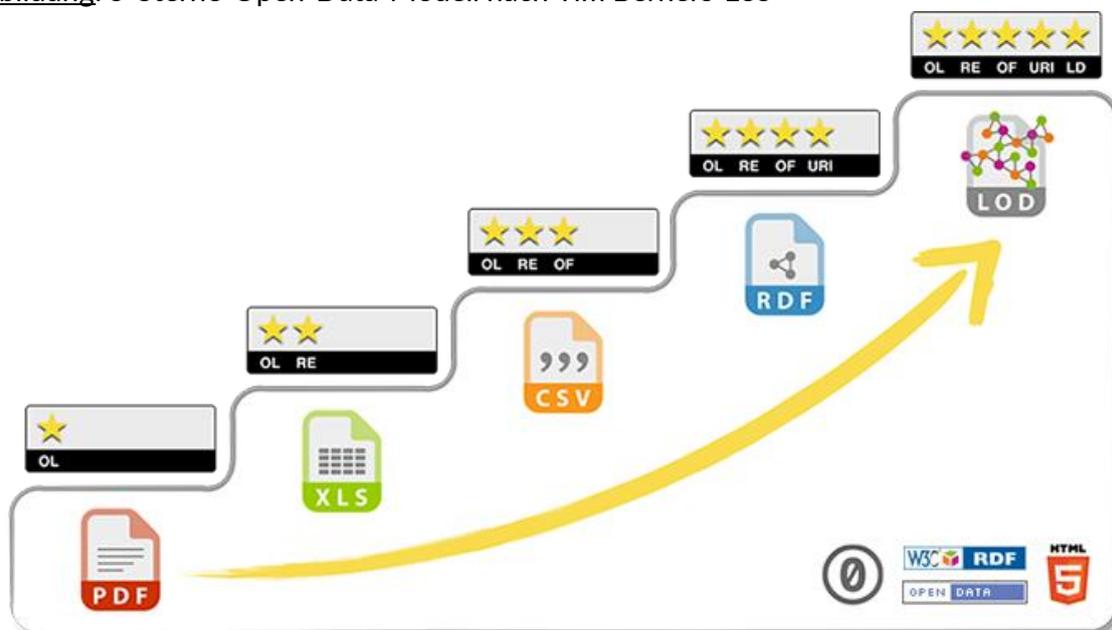
³ Vgl. Definition Linked Open Data bei Wikipedia: https://de.wikipedia.org/wiki/Linked_Open_Data. Tim Berners-Lee, Erfinder des World Wide Web, stellte 2009 in einem TED-Talk seine Zukunftsvision für ein künftiges WWW der Daten vor. Darauf aufbauend wurde das Modell der 5 Sterne für Open Data entwickelt: In den ersten Stufen sind Daten überhaupt im Web verfügbar, weitere Sterne gibt es für Daten in strukturierter Form (Tabellen). Die eigentlich visionäre Leistung Berners-Lees liegt jedoch in den Stufen 4 und 5 dieses Modells. Vier Sterne bedeutet, dass nicht nur eine bloße Tabelle vorliegt, sondern dass ein Datensatz sich selbst beschreibt: „Diese Spalte bildet Datum und Uhrzeit ab, in dieser Spalte befinden sich Adressen in einem definierten Format.“ Erst in der 5-Sterne-Ausbaustufe können verschiedene Datensätze so miteinander in Bezug gebracht – verlinkt – werden.

Wären die Daten in der RKI-Tabelle mit Metadaten selbstbeschreibend kodiert gewesen, hätten Veränderungen in der Tabelle automatisch ausgelesen werden können.

Dieses System mit wohlbeschriebenen Daten erlaubt die Zusammenfassung und Verknüpfung von Datensätzen aus gänzlich unterschiedlichen Quellen. Es ist somit weder die Einrichtung einer autoritativen und zentralen Datenplattform notwendig, noch müssen dafür zwingend die eingesetzten Fachverfahren beispielsweise über alle Berliner Bezirke hinweg vereinheitlicht werden (sie müssen aber reformiert und geöffnet werden). Im Gegenteil erlaubt erst das offene Bekenntnis zu Linked Open Data gleichzeitig eine hohe Zusammenfassbarkeit verschiedener Datenquellen sowie eine weitgehend freie Wahl der eingesetzten IT-Systeme. Diese Herangehensweise dürfte auch die Grundsätze des Föderalismus optimal berücksichtigen. Der geplante Data Hub ist demnach ein wichtiger und willkommener Zwischenschritt, um schon jetzt Daten beispielsweise aus Fachverfahren automatisiert zu exportieren und der Öffentlichkeit bereitzustellen. Wenn gleichzeitig jedoch bereits jetzt die strategischen Entscheidungen für die Bereitstellung von Linked Open Data getroffen werden, können künftig auch Daten aus verschiedensten Quellen berücksichtigt und gemeinsam verwendet werden, die sich aus praktischen Gründen nur mit unverhältnismäßigem Aufwand oder gar überhaupt nicht auf den Nenner eines überall durchgehend genutzten Systems bringen lassen.

Der Fokus sollte daher auf die Ziele Linked Open Data und automatisierte Bereitstellungsprozesse gelegt werden. Je besser das interne Datenmanagement organisiert ist, desto einfacher lassen sich auch komplexere Vorgänge wie Linked Open Data oder datengetriebene automatisierte Entscheidungsmechanismen umsetzen.

Abbildung: 5-Sterne-Open-Data-Modell nach Tim Berners-Lee



Quelle: <https://5stardata.info/de/>

Vom Ziel her denken und nicht verkomplizieren: Der Berlin Data Hub muss auf Linked Open Data ausgerichtet sein und sollte nicht überfrachtet werden

Das Nutzungsversprechen von Open Data kann nur eingelöst werden, wenn mit einem modernen Datenmanagementsystem gearbeitet wird. Auf dem geplanten Berlin Data Hub⁴ lasten daher schon jetzt hohe Erwartungen. **Der Data Hub ist ein unabdingbarer und wichtiger Baustein für das Berliner Datenmanagement, aber keine "eierlegende Wollmilchsau"**. Er wird nur gut funktionieren, wenn das gesamte Daten-Ökosystem (insb. IT-Infrastruktur, Workflows, Zuständigkeiten, Personal, Ressourcen, politische Führung, technische Umsetzung) gut aufgestellt ist. Es gilt daher, den Data Hub mit realistischen Erwartungen zu flankieren. Drei Prämissen sind bei der Umsetzungen maßgeblich.

Erstens, der Data Hub sollte von Anfang an darauf ausgerichtet sein, verknüpfte Datensätze zu hosten - auch wenn es aktuell davon noch sehr wenige Fälle gibt. Es wäre ein Fehler, den Data Hub als "Cloudspeicher für große Excellisten" zu verzwerger. Der Aufbau von Speicher- und Verbreitungssystemen für Linked Open Data (sog. Triple Stores⁵) muss strategisch im Vordergrund stehen. Die Grundprinzipien von Linked Data erlauben hierbei auch eine dezentrale Architektur mit mehreren, verteilten Triple Stores, die dennoch auch für gemeinsame Abfragen verwendet werden können.

Beispielhaft sei hier das auf Seite 3 beschriebene Problem genannt: Bisher wurden Faktendaten zum Beispiel in Word veraktet und ausgedruckt. Dritte, die diese Daten benötigen, müssen sie händisch aus den Akten suchen. Ein Lösungsansatz wäre hierbei, den bisherigen Pfad der Aktenerstellung mit Office-Produkten zu verlassen und die notwendigen Informationen textbasiert zu erfassen, d.h. Inhalt und Form voneinander zu trennen. Mit einem zu schaffenden Workflow kann dann einerseits automatisiert das angestrebte formatierte Dokument samt aller Notwendigkeiten (Briefkopf, Datum etc) erstellt und als E-Akte veraktet werden. Andererseits können die Informationen parallel zu semantisch angereicherten Daten verarbeitet werden, die in einem Triple Store abgelegt werden. Somit kann bereits die Infrastruktur aufgebaut werden, mit der interne Dritte die so abgelegten Daten effizienter abfragen können, ohne hierfür händisch Papierakten konsultieren zu müssen. Die Entscheidung, welche der so erfassten Daten auch als Open Data bereitgestellt werden sollen, ist danach eine reine Kategorisierungsfrage.

In der Vergangenheit wurden an mehreren Stellen bereits sogenannte ETL-Verfahren (Extraktion, Transformation, Laden) etabliert, mit denen die in Fachverfahren erhobenen Daten

⁴ Ein Konzept zur Entwicklung eines Berliner Data Hub wurde von KPMG 2019 erarbeitet; die [Abschlusspräsentation](#) mit den wichtigsten Ergebnissen ist hier verfügbar.

⁵ Ein Triple Store organisiert Informationen in einer Semantik mit den drei Bestandteilen „Subjekt – Prädikat – Objekt“ – deswegen Tripel. Als Beispiel kann das Wikidata-Datenobjekt für das Land Berlin herangezogen werden: „Berlin – ist Instanz von – Bundesland (Deutschland)“. So können beispielsweise Abfragen in der Form „zeige alle Bundesländer Deutschlands mit der zugehörigen Bevölkerungszahl in aufsteigender Reihenfolge an“ getätigt werden.

aus deren Datenbank exportiert und als Tabellen aufbereitet wurden, die danach über ein geeignetes Datenportal bereitgestellt werden. Diese Verfahren kamen bislang stets nur auf die Stufe von 3-Sterne-Daten, d.h. in Tabellen strukturierte Datensätze. Für die Weiterverwendung der Daten ist daher bislang stets eine händische Erstellung oder Anpassung von Auswertungsskripten notwendig. Um auch hier dem Ziel von 5-Sterne-Open-Data gerecht zu werden, ist es unerlässlich, diese ETL-Verfahren dahingehend anzupassen. Das heißt, dass die Semantik der Quelldatenbank durch die Verwendung standardisierter Vokabulare⁶ auch im Zieldatensatz kodiert und als Knowledge Graph⁷ abgelegt wird. Perspektivisch erscheint es auch sinnvoll, bei der Neubeschaffung von IT-Lösungen für Fachverfahren bereits deren interne Datenhaltung als Knowledge Graph zu fordern. Dies würde die bisherigen interimswise Extraktionsprozesse weitgehend ablösen, und bereits an der Quelle die Kompatibilität mit dem angestrebten Ziel sichern.

Zweitens, der geplante Data Hub soll kein monolithisches System sein, das alle erdenklichen Bedürfnisse und Anforderungen erfüllen kann. Vielmehr muss es sich um einen Verbund von Diensten und Systemen handeln, der verschiedene Aufgaben übernimmt. Hierbei kann modular vorgegangen werden, d.h. nicht alle Dienste und Systeme müssen am Anfang vorhanden sein. Sie können peu à peu dazukommen. So wächst der Data Hub in seinen Funktionalitäten und wird immer weiter verbessert. Folgende Aufgaben (teils grundverschieden) erscheinen sinnvoll:

- Auffindbarkeit und Katalogisierung
- Triple Store für Linked Data
- Speicherort für nicht-linked Data
- Zeitreihendatenbanken z.B. für Sensorikdaten
- Extraktions-, Transformations- und Ladeverfahren (ETL) für den automatisierten Import aus Datenbanken von Fachverfahren

Drittens, zur praktischen Umsetzung des Data Hub wurden in den Beteiligungsworkshops viele Hinweise und Ideen gegeben. Es ist daher erfolgversprechend, einen regelmäßigen Austausch zwischen Verwaltung und Dritten zu etablieren, um den Data Hub schrittweise aufzubauen und zu verbessern. Anekdotisch sind hier folgende Hinweise genannt:

- Daten sollten möglichst granular und disaggregiert vorliegen, sie sollten eine zeitliche und räumliche Dimension haben.

⁶ Die notwendige Standardisierungsarbeit für den Löwenanteil dieser Vokabulare ist in den vergangenen Jahrzehnten bereits durch internationale Standardisierungsgruppen wie das World Wide Web Consortium (W3C) durchgeführt worden. Ohne weitere Vorarbeit kann ein großer Teil der notwendigen Semantik durch die Verwendung der Vokabulare auf schema.org umgesetzt werden.

⁷ Die Gesamtheit der in einem Triple Store abgelegten Informationen wird als Knowledge Graph ("Wissensgraph") bezeichnet. Da die Informationen als Tripel kodiert und somit untereinander vernetzt sind, bilden sie gemeinsam ein Netzwerk bzw. einen Graphen. Verschiedene Knowledge Graphs können miteinander verbunden werden und bilden dann einen größeren gemeinsamen Graphen. Ein gutes Beispiel ist der Knowledge Graph hinter Wikidata. Eine Einführung in das Konzept ist hier zu finden: <https://towardsdatascience.com/a-very-simple-guide-of-knowledge-graph-for-health-data-scientists-bb8cefc1db31>.

- Datenformate sollten standardisiert vorgegeben werden; es sollten international gängige Lizenzen (z.B. Creative Commons) verwendet werden, um auch internationale Nutzung zu ermöglichen. Als Standard sollte die gemeinfrei-ähnliche Veröffentlichung unter CC-0 gewählt werden, zumal es in der Regel keinen urheberrechtlichen Schutz für Fakteninformationen gibt, der eine restriktivere Lizenz erlauben würde (z.B. mit Zwang zur Nennung der Quelle). Es ist wichtig, dass auch nicht tabellenbasierte Datensätze veröffentlicht werden können kann, insbesondere Video- und Bildformate (z.B. OpenParliamentTV).
- Es braucht viel mehr offene Schnittstellen zu den bestehenden Fachverfahren.
- Die Metadaten müssen deutlich verbessert werden; hier liegen massive Fehler, kaum ein Datensatz in Deutschland verfügt über „gute“ Metadaten (einfach, eindeutig, ohne Verwaltungssprache, nachvollziehbar; im Idealfall auch bilingual). Die Daten brauchen eine klare Identifikation sowie eine unverkennbare und unveränderbare ID sowie ein klares Vokabular (stabile, unveränderliche URIs).
- Der Data Hub sollte bilingual (deutsch und englisch) gestaltet werden, um den sowohl in Wissenschaft als auch Berliner Stadtgesellschaft stark vertretenen nicht-deutschen Muttersprachler:innen die Nutzung zu erleichtern.
- Upload- und Downloadprozesse im Data Hub sollten möglichst niedrigschwellig funktionieren (geringe Anforderungen an Datenformate, einfache Schnittstellen, etwaige Unterstützung). Der Download sollte auch ohne Limit funktionieren.
- Der Data Hub sollte die Möglichkeit haben, Datensätze zu kommentieren bzw. Verbesserungen zu melden.
- Zusätzlich zum Datenportal sollte es auch Visualisierungswerkzeuge (z.B. auch Dashboards) geben, mit denen Datensätze auch dargestellt und kontextualisiert werden können. Hier sind Opportunitätskosten zu berücksichtigen. Kann auf eine Linked-Open-Data-Infrastruktur zurückgegriffen werden, können Visualisierungen aufgrund der Verknüpfbarkeit der Daten mit wenig Aufwand und teilweise auch automatisiert erstellt werden. In den Vorstufen von Linked Open Data müssen diese Werkzeuge meist passgerecht angefertigt werden. Passende Dashboards können Verständnis und Akzeptanz für solide IT-Infrastruktur fördern, sollten aber nicht übermäßig mit den Ressourcen für den Aufbau der Linked-Data-Infrastruktur konkurrieren müssen.
- Der Data Hub sollte die Funktion haben, den Datennutzenden Stichworte oder Keywords zu empfehlen.
- Ziel des Data Hubs muss es auch sein, zwischen bestehenden Datenportalen verschiedener Ebene (Kommune, Land, Bund, EU) Daten kreuz-und-quer zu verlinken, Standards sind hier maßgeblich, bzw. als Dateninfrastruktur auch für zukünftige Projekte wie Urban Data Platform oder Digital Twins zu dienen.

Schlaglichter aus den Beteiligungsworkshops

- Der Nutzen von Open Data ist für viele Beteiligte, insbesondere aus Verwaltung und landeseigenen Betrieben, nicht nachvollziehbar. **Der Aufwand ist konkret vorhanden, der Nutzen nur abstrakt in der Zukunft versprochen.** Das Potenzial von Open Data

muss besser kommuniziert werden. Eine konsequent durchdachte Open-Data-Strategie führt zum Ausbau der in der Verwaltung ohnehin notwendigen Dateninfrastruktur samt der zugehörigen Prozesse. Diese haben das Potenzial, den Arbeitsalltag in der Verwaltung zu erleichtern und darüber auch den eigenen Nutzen direkt sichtbar zu machen.

- In der Verwaltung, aber auch darüber hinaus (Wissenschaft, Stadtgesellschaft, Wirtschaft), braucht es eine **“Open Data Mentality”**, um das Thema breiter zu verankern und voranzubringen.
- Es braucht **mehr politische Führung** beim Thema Open Data, besonders auf bezirklicher Ebene. Open-Data-Beauftragte haben teilweise unklare Zielstellungen und Mandate, es fehlt zudem an Ressourcen und Weiterbildungsmöglichkeiten. Es bedarf dazu, neben den erforderlichen IT-Ressourcen, auch einen **“Pioniergeist”** im Management, um die Datenöffnung und Datenteilung aktiv voranzutreiben.
- Die **Wirtschaft fordert den notwendigen Kulturwandel** nach dem Motto **“So kann es gehen!”**, dazu gehört auch ein aktives Experimentieren bei Open Data.
- **Datenmapping, Datenkooperation, Zuständigkeit für Öffnung:** Durch die Arbeits- und Organisationsweise der Verwaltung sind Zuweisungen von Zuständigkeit und Verantwortlichkeit zwischen allen Ebenen der behördlichen Verwaltungsstruktur für die Öffnung von Daten unumgänglich. In begrenztem Umfang kann auch die Erkundung vorhandener Datenbestände in der Verwaltung sinnvoll sein. Erkundung und Mapping dürfen jedoch kein Selbstzweck sein, sondern müssen unmittelbar von der Ertüchtigung der Dateninfrastrukturen hin zu Linked Open Data gefolgt werden. Gegebenenfalls ist die direkte Ertüchtigung sogar einem Mapping vorzuziehen (Opportunitätskostenprinzip).
- **Open Data wird nur funktionieren, wenn die bestehenden Fachverfahren Öffnung und Bereitstellung von Daten in die Arbeitsprozesse integrieren.** Fachverfahren auf allen Ebenen müssen daher überprüft und reformiert werden. Ein falscher Weg dagegen ist es, um die bestehenden Fachverfahren herum aufwändige Open Data Prozesse zu konstruieren (Doppelaufwand).
- Beispiele für Open Data und Linked Open Data sollten konkrete Anschauungen über **Effizienzgewinne für die Verwaltung, Kostenersparnisse und einfachere Zusammenarbeit** zwischen den Verwaltungen bieten. Akteur:innen aus der Zivilgesellschaft können helfen, diese ersten Beispiele zu erarbeiten.
- Das Berliner Energie- und Klimaschutzprogramm **BEK 2030⁸ ist ein gutes Beispiel für Daten** und sollte mehr für Auswertungen genutzt werden.

⁸ Vgl. Website BEK 2030: <https://www.berlin.de/sen/uvk/klimaschutz/klimaschutz-in-der-umsetzung/>.

- Die **Beantwortung parlamentarischer Anfragen** könnte durch passende Workflows begleitet werden, die eine Nachnutzung der hierfür herangezogenen Informationen in strukturierter Form ermöglichen.
- **Mehr Klasse als Masse**: Aktuell ist die Datenqualität der bereitgestellten Datensätze eher gering. Ziel sollte es sein, auf Qualität relevanter Datensätze zu setzen und in diesem Zuge die Automatisierung der Bereitstellung voranzutreiben, anstatt die Quantität minderwertiger und wenig genutzter Datensätze zu erhöhen.
- Ein **Bot-Tool** könnte dabei unterstützen, höhere Datenqualität zu erreichen, in dem der Bot Vorschläge macht, welche weiteren Maßnahmen es braucht, um die Daten in die nächste Sternekategorie von Linked Open Data zu heben.
- Wenn nur ein Bruchteil jener Ressourcen und Motivation, welche für die Umsetzung der DSGVO zur Verfügung standen, auch bei der Datenöffnung vorhanden wären, wären wir bei Open Data schon weiter. **Länder wie Großbritannien statten entsprechende Institutionen wesentlich und spürbar besser aus**. Der Ressourceneinsatz spiegelt hier auch einen politischen Willen wider.
- Mit der **Open Data Informationsstelle (ODIS)** verfügt Berlin über eine extrem wertvolle und nützliche Ressource, um das Thema in die Umsetzung zu bringen und dabei Wissensgrundlagen zu schaffen. Auch die landeseigenen Betriebe sollten die Möglichkeit erhalten, sich Beratung zur Umsetzung von Open Data einzuholen. Bzgl. Ressourcen der Hinweis auf das umfassende Open Data Handbuch Berlin⁹, welches auch als ein lebendiges Dokument auf github¹⁰ weitergeschrieben wird.
- Es besteht noch **Diskussions- und Erklärungsbedarf bei den Herausforderungen der landeseigenen Betriebe** in den Bereichen Rechtssicherheit, Datenschutz, Haftungsklausel, und der kritischen Infrastruktur - obwohl die meisten dieser Fragen bereits geklärt sind. Hier könnten Beratungsangebote hilfreich sein.

⁹ <https://daten.berlin.de/sites/default/files/berliner-open-data-handbuch.pdf>

¹⁰ <https://github.com/berlinonline/open-data-handbuch>

Ausblick

Mit der Beteiligung des Berliner Open Data Ökosystems im ersten Halbjahr 2022 ist sowohl eine umfassende Evaluation als auch eine Aufnahme der wichtigsten Bedarfe und Hinweise für die zukünftige Ausrichtung von Open Data in Berlin erfolgreich abgeschlossen. Es gilt nun, auf Basis dieser Ergebnisse und im weiteren Austausch mit alten und neuen Akteur:innen Open Data in Berlin neu aufzustellen. Dazu gehört die Formulierung von ambitionierten Zielen und die Konkretisierung von messbaren Handlungsempfehlungen für die nächsten Jahre. Dabei kann das Land Berlin auf die Erfolge der letzten Jahre wie zum Beispiel die Open Data Rechtsverordnung oder die Open Data Informationsstelle aufbauen. Open Data muss nun in die flächendeckende Umsetzung kommen, technisch und prozessual in die Abläufe integriert werden. Dazu ist von zentraler Stelle im Land Berlin strategisch, politisch, personell und prozessual zu flankieren. Linked Open Data ist dabei der zentrale Baustein für eine digitale, gemeinwohlorientierte und effiziente Verwaltung.