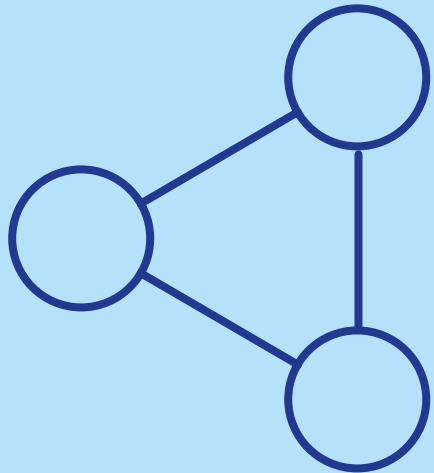


Studie

Linked Open Data in der Praxis

**Vernetzte Verwaltungsdaten am
Beispiel der Berliner Organigramme**



Eine Fallstudie der Open Data Informationsstelle Berlin
März 2025

Inhalt

Vorwort	4
Danksagung	6
Über die Autor:innen	7
Executive Summary	8
1. Einleitung:	
Eine solide Datenbasis für eine digitale Verwaltung	10
1.1. Dialog mit einem Berliner Bären	10
1.2. Die Daten bestimmen den Output	11
1.3. Berlins Strategie: LOD und ein Pilotprojekt	12
1.4. Über diese Studie	14
2. Das Fundament:	
Linked Data im Überblick	16
2.1. Was ist Linked Data?	16
2.2. Tim Berners-Lees und der Weg zu den 5-Sternen	16
2.3. Die Rolle von Linked Data in der Praxis	18
2.4. LOD-Pilotprojekte im Verwaltungskontext	19
3. Die Fallstudie:	
Organigramme als Linked Open Data	22
3.1. Die Ausgangslage: Organigramme als Paradebeispiel für unstrukturierte Informationssilos	22
3.2. Die Vision: Das Wissensnetzwerk der Berliner Verwaltungshierarchie	25
3.3. Die Lösung: Das Organigramm-Tool als digitales Hilfsmittel	27
3.4. Linked Open Data erstellen: Vom PDF-Organigramm zum Triple Store	30

3.4.1	Prinzip 1: Das Organigramm-Tool als URI-Generator und Editor	31
3.4.1	Prinzip 2: Organigramm-Daten als Triples	33
3.4.1	Prinzip 3: Eine Ontologie speziell für Berliner Organigramme	35
3.4.1	Prinzip 4: Organigramme mit Daten im Web verlinken	37
3.5.	Linked Open Data nutzen: Die Organigramme befragen	39
3.5.1	Anwendungsfall 1: Den Organigramm-Triple Store als Wissensspeicher über die Berliner Verwaltung nutzen	41
3.5.1	Anwendungsfall 2: Organigramme mit externem Wissen verknüpfen	46
3.9.	Ausblick: Linked Data leichter nutzen mit KI	48
4.	Erkenntnisse und Handlungsempfehlungen	48
4.1.	Kompetenzen: Notwendiger Perspektivwechsel von Menschen- zu Maschinenzentriertheit	48
4.2.	Standards: Organigramme als unerwartet komplexes Fallbeispiel	50
4.3.	Infrastruktur: Lösungen über den Prototypenstatus hinaus benötigt	52
4.4.	Kritische Masse und Vernetzung: Wir brauchen mehr Linked Data	53
5.	Ausblick und Fazit	56
	Abbildungsverzeichnis	60
	Glossar wichtiger Begriffe	62
	Weiterführende Literatur und Leseempfehlungen	65
	Impressum	68

Vorwort

Smart Cities, die Entwicklung von Machine Learning und KI-gestützten Anwendungen prägen die digitale Gegenwart. Um das Vertrauen der Menschen in staatliche Institutionen zu stärken, braucht es transparentes Verwaltungshandeln. Die Grundlage dafür: verlässliche, vernetzte, offene und vor allem maschinenlesbare Daten. Genau hier setzt das Konzept von Linked Open Data (LOD) an. Bereits in den frühen 2000er Jahren entwarf Tim Berners-Lee die Vision eines Netzes aus verknüpften Informationen. Doch während diese Vision in vielen Bereichen Realität wurde, ist sie in der öffentlichen Verwaltung noch weitgehend unbekannt. Diese Studie verfolgt einen praxisnahen Ansatz: Am Beispiel der Berliner Verwaltungsorganigramme wird der Weg von statischen PDFs hin zu vollständig vernetzten, maschinenlesbaren Informationen auf der höchsten Stufe des 5-Sterne-Modells des Datenreifegrads aufgezeigt. Die Erkenntnisse sind vielfältig. Besonders wichtig: Es

bedarf einer Perspektivänderung dahin, Daten für Maschinen und nicht allein für Menschen lesbar aufzubereiten. Für eine erfolgreiche Umsetzungen sind behördenübergreifende Kollaboration, engagierte Pionier:innen, Datenkompetenz sowie die nötige technische Infrastruktur unverzichtbar. Gerade in dieser technischen Herangehensweise liegt ein tieferer Wert für das Leitbild einer menschengerechten Digitalisierung: Wenn Daten so bereitgestellt werden, dass Maschinen sie optimal verarbeiten können, entsteht die Grundlage für Anwendungen, die den Menschen tatsächlich dienen – mit größerer Zuverlässigkeit, Transparenz und Zugänglichkeit.

Die vorliegende Studie markiert einen Schritt auf diesem Weg. Ich wünsche eine anregende Lektüre und viele gewinnbringende Erkenntnisse.

Nicolas Zimmer

Vorstandsvorsitzender der Technologiestiftung Berlin

Danksagung

Wir bedanken uns bei all den Daten-Enthusiast:innen, die sich mit auf die Reise zum 5-Sterne-Reifegrad – dem Ideal für offene, vernetzte und maschinenlesbare Verwaltungsdaten – gemacht haben. Durch wertvolles Fachwissen, Feedback, Mitarbeit in Workshops, dem Anlegen von Daten und dem Gewinnen neuer Supporter:innen haben sie dieses Projekt aktiv mitgestaltet.

Ein besonderer Dank gilt den behördlichen Beauftragten für Open Data des Landes Berlin, insbesondere Julia Schabos und William Treherne, sowie der zentralen Verantwortlichen für Open Data im Land Berlin Betül Özdemir, Knud Möller von BerlinOnline und dem engagierten Open Data Team der Open Knowledge Foundation.

Über die Autor:innen

Seit 2018 begleiten wir als Open Data Informationsstelle Berlin (ODIS) die Stadt auf dem Weg zu einer partizipativen, nachhaltigen und datengetriebenen Gesellschaft. Wir sehen uns dabei als Begleiterin und Unterstützerin der Berliner Verwaltung und als Schnittstelle zur Stadtgesellschaft. Wir unterstützen bei der Veröffentlichung von Daten, vernetzen Akteur:innen des Open-Data-Ökosystems, entwickeln digitale Anwendungen mit Open Data und formulieren Handlungsempfehlungen für zukünftige Maßnahmen zur Förderung offener Daten und für ein besseres Datenmanagement.

Die ODIS ist ein Projekt der Technologiestiftung Berlin. Gefördert von der Senatskanzlei Berlin und der Investitionsbank Berlin aus den Mitteln des Landes Berlin. Mehr Informationen und aktuelle Projekte finden sich online unter odis-berlin.de.

Executive Summary

Anfang der 2000er Jahre entwickelte Tim Berners-Lee, der Erfinder des World Wide Web, das Konzept von Linked Data sowie ein 5-Sterne-Modell zur Bewertung der Datenqualität. Ziel war es, Daten maschinenlesbar zu machen und deren Verarbeitung durch Algorithmen und Programme zu optimieren.

Heute, knapp 20 Jahre später, eröffnen neue Technologien wie Künstliche Intelligenz enorme Möglichkeiten zur Analyse und Verknüpfung von Daten. Voraussetzung dafür ist ein hoher Reifegrad der zugrunde liegenden Daten. In der Berliner Verwaltung zeigt sich jedoch ein anderes Bild: Viele Informationen werden weiterhin primär für die menschliche Lesbarkeit statt für eine maschinelle Verarbeitung aufbereitet. Dadurch bleibt das Potenzial ungenutzt, Verwaltungsdaten effizient auszuwerten und innovative Anwendungen zu ermöglichen.

Unsere Studie zeigt: Linked Open Data (LOD) kann dieses Potenzial heben, erfordert aber gezielte Anstrengungen und ein technisches und organisatorisches Fundament. In einem über zwei Jahre andauerndem Projekt der Open Data Informationsstelle Berlin haben wir untersucht, wie LOD in der Verwaltung angewendet werden kann. Unser Ziel war es, den Schritt von der Theorie in die Praxis zu machen und Herausforderungen auf dem Weg zum höchsten Reifegrad von Verwaltungsdaten zu identifizieren. Wir wollten außerdem verstehen, welche Anwendungsfälle sich aus LOD ergeben.

Dafür haben wir ein praktisches Beispiel gewählt: die Berliner Verwaltungsorganigramme. Diese lagen bislang nur als statische PDF-Dateien auf den verschiedenen Webauftritten der Behörden vor – ein klassisches Informationssilo. Um sie maschinenlesbar bereitzustellen und zu einem Wissensnetzwerk zu verknüpfen, haben wir ein Tool entwickelt, das die Prinzipien von LOD berücksichtigt und eine Bereitstellung auf dem höchsten Reifegrad ermöglicht. Mehrere Berliner Behörden konnten so ihre ersten LOD-Organigramme erstellen und veröffentlichen.

Während des Projekts haben wir wertvolle Erkenntnisse gewonnen und konkrete Handlungsempfehlungen abgeleitet, um LOD in der öffentlichen Verwaltung Berlins zu etablieren. Dabei wurde klar: Die Umsetzung von LOD erfordert nicht nur technische, sondern auch organisatorische und kulturelle Veränderungen. Neben einer geeigneten technischen Infrastruktur, etwa einem zentralen Triple Store und Ontologie-Repositories, ist vor allem ein Perspektivwechsel hin zu maschinenlesbaren Daten entscheidend. Dafür ist die gezielte Akquise von Fachpersonal mit Datenexpertise in der Verwaltung und eine Kultur des Datenteilens nötig.

Darüber hinaus könnte Künstliche Intelligenz einen großen Auftrieb für LOD geben, insbesondere in der Datenanalyse und -verarbeitung. Die Integration dieser Technologien könnte neue, innovative Lösungen für die Verwaltung ermöglichen. Es bleibt jedoch noch zu erforschen, wie das Zusammenspiel von LOD und KI das volle Potenzial entfalten kann.

Die Etablierung von LOD ist jedoch kein einmaliges Projekt, sondern ein fortlaufender Prozess. Unser Praxisbeispiel zeigt: LOD muss schrittweise aufgebaut und stetig weiterentwickelt werden. Ein einzelnes Vorhaben reicht nicht aus – es bedarf einer langfristigen Strategie, um offene Verwaltungsdaten mit hoher Qualität und Vernetzung bereitzustellen.

Der Ausblick bleibt dennoch positiv: Aufbauend auf unserer Vorarbeit können zahlreiche weitere Fachdaten veröffentlicht und vernetzt werden. Erst dann wird das volle Potenzial von Linked Open Data für Verwaltung, Wirtschaft und Gesellschaft wirklich greifbar.

1. Einleitung:

Eine solide Datenbasis für eine digitale Verwaltung

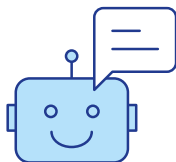
Für die öffentliche Verwaltung haben Daten unverkennbar eine zentrale Rolle. Sie werden nicht nur in großen Mengen tagtäglich erhoben und bilden das Fundament des Verwaltungshandelns, sondern ermöglichen fundierte Entscheidungen und transparente, bürgernahe Kommunikation.

Mit der fortschreitenden Digitalisierung entfalten sich weitere Möglichkeiten. Technologien wie Künstliche Intelligenz, maschinelles Lernen und Big Data Analytics versprechen bahnbrechende Innovationen, die den Umgang mit Informationen, effiziente Arbeitsprozesse und Services auf gewinnbringende Weise transformieren werden – doch damit gehen auch zahlreiche Anforderungen an die Daten einher. Wie ist die Verwaltung auf diesen Wandel vorbereitet?

1.1. Dialog mit einem Berliner Bären

Ein Blick zurück ins Frühjahr 2023: Der Krimi um das Berliner Wahldebakel rückt langsam aus dem Rampenlicht der öffentlichen Berichterstattung. Die darauffolgenden Wiederholungswahlen und der Regierungswechsel sind fast schon Geschichte, als eine Meldung für den ein oder anderen Schmunzler sorgt: Wie die Presse berichtet, scheint der Berliner Chatbot des Service-Portals, verkörpert durch einen Bären namens *Bobbi*, auch mehrere Wochen nach der Wahl noch nicht auf dem neuesten Stand der politischen Entwicklungen zu sein. Trotz des Wechsels im Amt des Regierenden Bürgermeisters zu Kai Wegner behauptet Bobbi felsenfest, Franziska Giffey sei Inhaberin dieses Amtes. In Wahrheit ist sie nicht mehr Regierende Bürgermeisterin sondern „nur“ noch Bürgermeisterin von Berlin, also Wegners Stellvertreterin. Zur gleichen Zeit und augenscheinlich im Kontrast stehend, diskutiert gefühlt die gesamte Technologiebranche über die massiven

Potenziale durch die rasanten Entwicklungen im Bereich der künstlichen Intelligenz, angeregt durch Sprachmodelle wie ChatGPT.



Also Bobbi, wer ist Regierender Bürgermeister in Berlin? „Die Regierende Bürgermeisterin von Berlin ist Franziska Giffey.“

Ähh... dann vielleicht so: Bobbi, wer ist Kai Wegner? „Es tut mir sehr leid. Leider habe ich zu Ihrer Frage keine passende Antwort gefunden.“¹

1.2. Die Daten bestimmen den Output

Genaugenommen wurde Bobbi nicht dafür programmiert, Fragen zu Verwaltungsstrukturen zu beantworten, sondern war ausschließlich als Hilfestellung rundum das Service-Portal² gedacht. Diese Anekdote illustriert aber gerade deshalb einen Aspekt sehr anschaulich: Jeder Chatbot, ebenso wie jedes andere Softwareprogramm, jedes Webtool und jede Datenanalyse, sind zum einen zweckgebunden und werden für spezifische Aufgaben entwickelt. Zum anderen sind sie auch nur so leistungsfähig und aussagekräftig, wie die Informationen und Daten, mit denen sie entsprechend des Use Cases gespeist werden. Ohne eine solide Datengrundlage bleibt der Zauber der Technologie reine Illusion. In diesem Kontext steht Bobbi symbolisch daher nicht für geringen Innovationscharakter, sondern viel mehr für die Hürden, mit denen digitale Tools beim Thema Nutzendenführung, Relevanz und Genauigkeit konfrontiert sein können.

Mit zunehmender Digitalisierung stehen Verwaltungen daher vor einer zentralen Herausforderung: Gute Datenqualität, Datenaktualität und

1 Tagesspiegel Checkpoint – Künstliche Impertinenz: Berlins Chatbot Bobbi kennt Kai Wegner nicht (10.05.2023)

2 Bobbi war ein virtueller Bürger-Assistent im ServicePortal Berlin, der Nutzenden half, relevante Verwaltungsinformationen durch natürliche Sprachverarbeitung und maschinelles Lernen zu finden. Aufgrund veränderter Nutzendenerwartungen und der Weiterentwicklung moderner KI-Chatbots lief der Betrieb zum 31.12.2024 aus, wobei die gewonnenen Erkenntnisse in zukünftige KI-Anwendungen einfließen sollen: <https://service.berlin.de/chatbot/>

effektives Datenmanagement sind keine optionalen Aufgaben mehr, sondern essenzielle Grundlagen für Weiterentwicklung und Innovation. Ein modernes Datenmanagement, welches sich am FAIR-Prinzip³ orientiert, darf sich daher künftig nicht mehr nur auf die Erhebung und Bereitstellung unstrukturierter Informationen oder isolierter Datenbanken beschränken, da dies Innovation hemmt.

Die These: Wenn Daten künftig besser standardisiert, kontextualisiert und miteinander verknüpft werden können, schaffen sie die Grundlage für Digitalisierung und innovative Anwendungen. Das Konzept von Linked Open Data (LOD) stellt genau diese Qualitäten der Daten sicher.

1.3. Berlins Strategie: LOD als Maßnahme und ein Pilotprojekt

Berlin hat die Notwendigkeit zur Fokussierung auf Datenqualität im Kontext der Verwaltungsdigitalisierung erkannt. So wurde mit der 2023 verabschiedeten Open Data Strategie des Landes Berlin LOD als Entwicklungsziel ausgerufen, um wichtige Daten aus dem Verwaltungskontext künftig in qualitativ hochwertiger, zuverlässiger Form für digitale Anwendungen zur Verfügung zu stellen. Zu diesem Zeitpunkt existierte in der Berliner Verwaltung nahezu keine Erfahrung in der Erstellung und Veröffentlichung von Daten in diesem Format.

Für uns als Open Data Informationsstelle Berlin (ODIS), sind Prototypen ein bewährtes Mittel, um neue Entwicklungen konkret in der Praxis zu erproben. Mit ihnen können wir am „lebenden Objekt“

3 Die internationalen FAIR-Prinzipien sind Leitlinien für die Beschreibung, Speicherung und Veröffentlichung von u.a. Verwaltungsdaten. FAIR steht für: Findable – auffindbar, Accessible – zugänglich, Interoperable – interoperabel, Reusable – wiederverwendbar.

Erfahrungen machen, Hürden in der Implementierung aufdecken, Diskussionen anstoßen und Handlungsempfehlungen ableiten.

Daher haben wir im Rahmen eines Pilotprojekts ein prototypisches Tool entwickelt, mit welchem wir an einem ganz konkreten Use Case, der Erstellung von maschinenlesbaren Organigrammen, das Konzept von LOD Schritt für Schritt erproben können. Ganz nebenbei wurden gemeinsam mit den behördlichen Beauftragten für Open Data im Land Berlin im Jahr 2024 Standards für den Aufbau von Organigrammen, sowie Rollen und Positionen entwickelt.

Exkurs: Linked Open Data in der Berliner Open Data Strategie

Die Entwicklung der Berliner Open Data Strategie⁴ erfolgte in einem Partizipationsprozess. Vertreter:innen aus Verwaltung, Wirtschaft, Wissenschaft und Zivilgesellschaft brachten über eine Online-Umfrage und in Stakeholder-Workshops ihre Perspektiven, Anregungen und Impulse ein. Die Ergebnisse wurden in konkrete strategische Maßnahmen übersetzt. Die Maßnahme „Qualität & Technische Interoperabilität mit Linked Open Data“ verankert auch Aspekte zur Schaffung von technischer Infrastruktur und zur Modernisierung von Fachverfahren. Das Thema Linked Data wurde insbesondere durch Vertreter:innen aus der Zivilgesellschaft in die Strategie eingebracht.

4 Open Data Strategie Berlins von 2023: <https://www.berlin.de/moderne-verwaltung/e-government/open-data/strategieprozess/artikel.1257333.php>

1.4. Über diese Studie

Diese Studie bietet nicht nur eine fundierte Einführung in die Prinzipien von LOD, sondern veranschaulicht das Konzept anhand konkreter Projekterfahrungen und der identifizierten Herausforderungen.

Die Studie richtet sich dabei an ein breites Publikum: Entscheidungstragende in Politik und Verwaltung, die die Bedeutung vernetzter Daten für strategische Entscheidungen erkennen, und Fachleute, die LOD-Projekte planen und umsetzen. Sie ist besonders relevant für Personen, die in Städten, Kommunen oder im Bund tätig sind und an der Vernetzung von Verwaltungsdaten interessiert sind. Auch für andere Verwaltungen, die ähnliche Open Data Projekte umsetzen oder das Organigramm-Tool nutzen möchten, bietet die Studie wertvolle Einblicke. Ebenso finden Personen, die innovative Anwendungen für offene Verwaltungsdaten entwickeln oder bestehende Projekte erweitern möchten, hier wertvolle Impulse.

Ziel dieser Studie: Wir widmen uns der Frage wie in einer Stadt wie Berlin ein Datennetz von Grund auf entstehen kann. Sie gibt konkrete Handlungsempfehlungen zur Etablierung von LOD und zeigt auf, welche Potenziale sich daraus ergeben. Gleichzeitig soll sie den Dialog über LOD fördern und Synergien mit bestehenden, sowie geplanten Projekten anregen.

Das Kapitel 2 dient dazu einen Überblick über die Entstehung und den aktuellen Stand von LOD zu vermitteln.

Das Kapitel 3.1 widmet sich ganz konkret der Bedeutung von Datenqualität und vernetzten Daten an unserem Fallbeispiel der Berliner Verwaltungsorganigramme.

Kapitel 3.2 zeigt anhand des prototypischen Organigramm-Tools detailliert auf, wie die vier Grundprinzipien von Linked Open Data in der Praxis umgesetzt wurden.

Das Kapitel 3.3 zeigt anhand der im Fallbeispiel erstellten LOD-Organigramme exemplarisch auf, welche Mehrwerte durch die intelligente Verbindung verschiedener Datenquellen entstehen können.

Im Anschluss fasst Kapitel 4 zentrale Erkenntnisse zusammen und leitet konkrete Handlungsempfehlungen ab.

Abschließend werden in Kapitel 5 zukünftige Entwicklungen und Potenziale für den weiteren Einsatz von LOD im Land Berlin anskizziert.

2. Das Fundament: Linked Data im Überblick

Bevor wir in die konkrete Fallstudie einsteigen, geben wir im folgenden Kapitel einen Überblick über das Konzept von Linked Data, seine Genese und den Status quo zur bisherigen Anwendung in der Praxis.

2.1. Was ist Linked Data?

Linked Data ermöglicht es, Daten so zu strukturieren und zu veröffentlichen, dass sie aus verschiedenen Quellen miteinander verbunden und für verschiedene Anwendungen genutzt werden können. Anstatt Datensätze isoliert in unterschiedlichen Datenbanken oder Cloud-Speichern zu belassen, sorgt Linked Data dafür, dass Datensätze über die Grenzen dieser Systeme hinaus miteinander „sprechen“ können. Die Idee dahinter ist, dass dezentral vorliegende, aber miteinander verknüpfte Informationen, eine intelligentere und flexiblere Nutzung ermöglichen. Über ein solches Netz an Informationen können Systeme und Anwendungen auf vielfältige Weise miteinander interagieren. Dies macht Linked Data zu einem wichtigen Konzept für die Zukunft von Datenverarbeitung und Wissensmanagement.

Linked Open Data (LOD) erweitert das Konzept von Linked Data um den Aspekt, dass die Daten öffentlich zugänglich und frei nutzbar sind.

2.2. Tim Berners-Lee und der Weg zu den 5-Sternen

Die Wurzeln von Linked Data liegen in den Grundprinzipien des World Wide Web und wurden von dessen Begründer Tim Berners-Lee in den frühen 1990er-Jahren entwickelt. Ursprünglich konzentrierte sich das Web auf die Verlinkung von Dokumenten durch Hyperlinks. Doch Berners-Lee erkannte schnell, dass diese Idee auch auf Daten ausgedehnt werden könnte. Sein Ziel war es, Informationen im Internet nicht nur zugänglich zu machen, sondern sie auch maschinenlesbar und kontextualisierbar zu gestalten. Mit der Einführung des zugehörigen

Begriffs „Semantic Web“ im Jahr 2001 begann die theoretische Fundierung von Linked Data.⁵

2006 konkretisierte Berners-Lee seine Vision mit dem Konzept von Linked Data⁶. Er definierte Prinzipien, die bis heute die Grundlage bilden: Ressourcen müssen mit eindeutigen Identifikatoren (URIs) identifiziert werden, diese müssen über das Web abgerufen werden können, die bereitgestellten Daten sollten in maschinenlesbaren Standards wie RDF (Resource Description Framework) vorliegen und zu guter Letzt sollen die Daten Links zu anderen Datenquellen bzw. deren URIs enthalten. So können die Daten miteinander verknüpft und zu Netzen aus Informationen verbunden werden, auch bezeichnet als Wissensgraphen, den „Knowledge Graphs“. Was genau hinter diesen Prinzipien steckt, erläutern wir anhand der Fallstudie in Kapitel 3.2.

Ein weiterer Meilenstein war die Einführung von Berners-Lees 5-Sterne-Modell für Datenqualität. Dieses Modell beschreibt den sogenannten Reifegrad von Daten, und ist eine Qualitätseinschätzung hinsichtlich ihrer Bereitstellung. Das Modell gibt somit vor, wie sich offene Daten schrittweise verbessern lassen – von einfacher Online-Verfügbarkeit von Textdokumenten als PDF-Datei (ein Stern) über die Bereitstellung in maschinenlesbaren Formaten (drei Sterne) bis hin zur vollständigen Integration in das Semantic Web (fünf Sterne). Die höchsten Stufen des Modells setzen auf alle Prinzipien von Linked Data.

Das Modell der unterschiedlichen Reifegrade impliziert auch: Ob und wie offene Daten tatsächlich genutzt werden, hängt maßgeblich vom Reifegrad der Bereitstellung ab, weshalb dieser auch als Erfolgskriterium für die Umsetzung von Open Data generell herangezogen wird.⁷

5 Vgl. Berners-Lee, T., Hendler, J., Lassila, O., 2001: The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities. Scientific American.

6 Blogartikel „Linked Data“ von Tim Berners-Lee, 2006: <https://www.w3.org/DesignIssues/LinkedData.html>

7 Die Open Knowledge Foundation hat 2024 erstmals ein Open Data Ranking erstellt, das eine Bewertung der Offenheit von Daten auch über qualitative Parameter zulässt: <https://opendata-ranking.de>

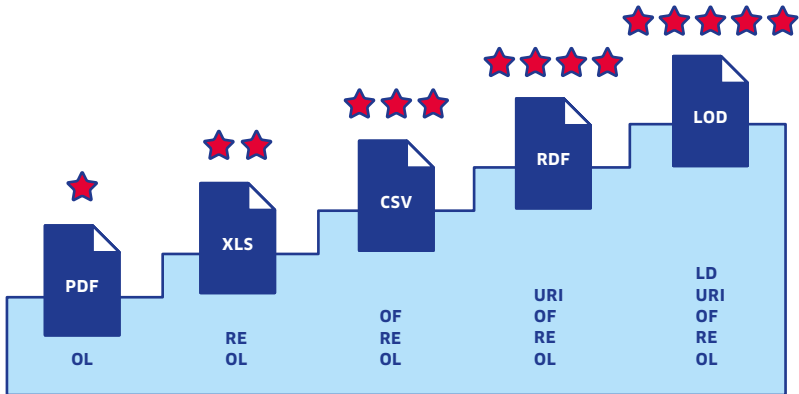


Abb. 1 Das Datenreifegrad-Modell nach Tim Berners-Lee.

2.3. Die Rolle von Linked Data in der Praxis

Seit der Einführung von Linked Data hat sich die Technologie weiterentwickelt und in der zunehmend datengetriebenen Welt praktische Anwendungen gefunden. Insbesondere die Entwicklung von Wissensgraphen für innovative Geschäftsmodelle von Unternehmen, aber auch in einigen Forschungsbereichen wird auf die Potenziale von verknüpften Daten gesetzt.

Suchmaschinen wie Google verwenden Wissensgraphen, um präzisere Suchergebnisse zu liefern, indem sie Informationen aus verschiedenen Quellen miteinander verbinden. Im E-Commerce setzen Unternehmen wie Amazon Linked Data ein, um personalisierte Empfehlungen zu erstellen und Produktinformationen zu verknüpfen.

Im wissenschaftlichen Kontext wird Linked Data schon lange in Bereichen eingesetzt, wo es um das Strukturieren von Informationen von großen Sammlungen geht, etwa im Bibliothekswesen, aber auch im kulturellen Kontext. Ein prominentes, öffentliches Beispiel ist die

Wissensdatenbank Wikidata⁸. Wikidata wurde entwickelt, um Wikipedia und weitere Projekte zu unterstützen, indem es eine zentrale Quelle für strukturierte Daten über Personen, Orte, Organisationen und vieles mehr bietet. In Kapitel 3.2 demonstrieren wir die Verknüpfung von Verwaltungsdaten mit Daten von Wikidata.

In der Verwaltungslandschaft ist das Konzept von Linked Data bisher wenig zur Anwendung gekommen. Eine Herausforderung dürfte dabei in der Integration von veralteten Systemen, der Bereitstellung von qualitativ hochwertigen und verknüpften Daten sowie der Sicherstellung der Interoperabilität zwischen verschiedenen Verwaltungseinheiten und Datensystemen liegen. Im europäischen Raum stellen Datenportale zunehmend Metadaten (also Angaben zu Titel, Beschreibung, Zuständigkeit etc.) als LOD bereit.⁹ Das Veröffentlichen der Daten selbst nach LOD-Prinzipien beschränkt sich dagegen auf wenige Pilotprojekte.

2.4. LOD-Pilotprojekte im Verwaltungskontext

Vorreiterin in Sachen LOD der öffentlichen Verwaltung im deutschsprachigen Raum ist die Schweiz. Das Schweizerische Bundesarchiv hat bereits 2017 den Linked Data Service LINDAS gestartet.¹⁰ Darüber können Schweizer Behörden ihre Daten in Form von Knowledge Graphs zur Verfügung stellen. Dies kann sowohl für offene Behördendaten, die sich an die Allgemeinheit richten, als auch für verwaltungsinterne Daten geschehen. Zum Stand Ende 2024 sind über 130 Datensätze auf der Website von LINDAS verzeichnet. Die Umwandlung und Veröffentlichung von Daten stellt Behörden jedoch teils vor Herausforderungen bezüglich des Verhältnisses von Aufwand und Nutzen und trifft nicht

8 Wikidata enthält derzeit über 110 Millionen Datensätze die nach Linked Open Data Prinzipien vorliegen: https://www.wikidata.org/wiki/Wikidata:Data_access/de

9 Metadaten als Linked Data werden z.B. durch das Datenportal der Europäischen Union (<https://data.europa.eu/de>) und durch das zentrale Datenportal des Bundes GovData (<https://www.govdata.de>) bereitgestellt.

10 LINDAS: <https://lindas.admin.ch/ecosystem/about-LINDAS>

immer den Bedarf, wie der Bericht einer Arbeitsgruppe des Kantons Zürich zur Erprobung von Linked Data darlegt.¹¹

Ebenfalls aus der Schweiz ist das Projekt LinkedDataGPT.¹² Die Webagentur Liip hat im Jahr 2023 einen Proof of Concept entwickelt, um die Abfrage von Linked Data der Stadt Zürich mittels Künstlicher Intelligenz zu ermöglichen. Auf Basis der ChatGPT-API können Nutzende im Tool in natürlicher Sprache Fragen an die Daten stellen. Diese werden in programmatischen Abfragen in die Sprache SPARQL umgewandelt, die wiederum die verlinkte Datenbasis abfragen. So können auch Personen ohne Programmier- und Linked Data-Kenntnisse mit den Daten interagieren.

In Berlin wurden bis zum Zeitpunkt Ende 2024 keine offiziellen Linked Data der Haupt- und Bezirksverwaltung veröffentlicht. Es existiert allerdings ein öffentlich verfügbarer erster Linked Data Datensatz, der vom landeseigenen Webdienstleister BerlinOnline bereitgestellt wird.¹³ Es handelt sich dabei um die Lebensweltlich orientierten Räume (LORs) Berlins, geografische Untereinheiten, die als Grundlage für Planung, Prognose und Beobachtung demografischer und sozialer Entwicklungen dienen. Die Basis dafür bildet ein strukturierter offener Datensatz in einem Geodatenformat, der in das Linked Data-fähige RDF-Format umgewandelt wird.

Weiterhin hat sich im Rahmen des Vierten Nationalen Aktionsplanes¹⁴ die Berliner Senatsverwaltung für Finanzen verpflichtet, den bereits seit vielen Jahren als offene Daten veröffentlichten Doppelhaushalt als Linked Data im RDF-Format zu veröffentlichen. Mit tausenden

11 Linked Open Data, Abschlussbericht der Erprobungsphase des Kantons Zürich: https://www.zh.ch/content/dam/zhweb/bilder-dokumente/themen/politik-staat/offene-behoerendendaten-nutzen/Abschlussbericht%20LOD%20Erprobung_public.pdf

12 LinkedDataGPT: <https://ld.gpt.liip.ch>

13 Berliner LORs als Linked Data bereitgestellt von BerlinOnline: <https://berlinonline.github.io/lod-berlin-lor>

14 Vierter Nationaler Aktionsplan von Open Government Deutschland: <https://www.open-government-deutschland.de/opengov-de/ogp/aktionsplaene-und-berichte/4-nap/berliner-haushaltsdaten-als-linked-open-data-verpflichtung-berlin--2225466#tar-2>

von Ausgabe- und Einnahmetiteln, also Informationen zur Höhe und Zweck von Einnahmen und Ausgaben, enthält der Haushalt wichtige Informationen mit örtlichem und politischem Bezug. Dieses Pilotprojekt wird gemeinsam mit der Staatskanzlei Schleswig-Holstein umgesetzt, eine erste Veröffentlichung der Haushaltsdaten als Linked Data ist für 2025 geplant.

3. Die Fallstudie: Organigramme als Linked Open Data

In diesem Kapitel beschreiben wir, wie wir uns dem Potenzial von LOD nicht nur konzeptionell, sondern auch praktisch genähert haben. Dafür haben wir uns auf einen speziellen Use Case fokussiert: Die Erstellung von LOD-Organigrammen.

3.1. Die Ausgangslage: Organigramme als Paradebeispiel für unstrukturierte Informationssilos

Die Berliner Verwaltung ist vielschichtig – ebenso wie die Art, wie sie Daten erhebt, verarbeitet und teilt. Während in einigen Sektoren wie Finanzen und Geodaten bereits überwiegend mit standardisierten, maschinenlesbaren Daten gearbeitet wird, fehlen in anderen Bereichen einheitliche Vorgaben. Teils werden Daten bisher gar nicht maschinenlesbar bereitgestellt. Ein Beispiel dafür sind Informationen die klassischerweise als Organigramm abgebildet werden.

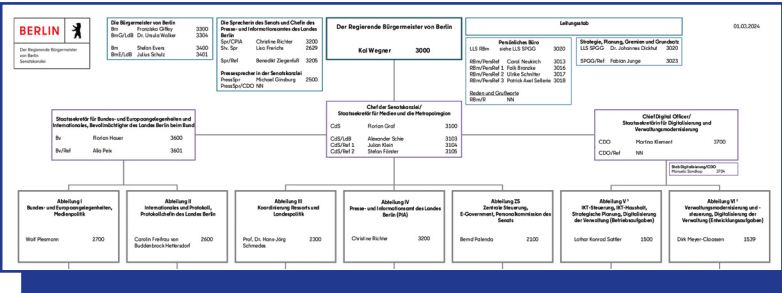


Abb. 2 Exemplarischer Ausschnitt aus dem PDF-Organigramm der Berliner Senatskanzlei (Stand: März 2024)

Ein Organigramm ist eine grafische Darstellung des Aufbaus einer Organisation, die organisatorische Einheiten, Aufgabenverteilung (bzw. Hierarchieebenen) und Kommunikationswege offenlegt. Im Kontext der öffentlichen Verwaltung stellen Organigramme wichtige Informationsquellen über die verschiedenen Verwaltungseinheiten und ihre Beziehungen zueinander dar, die vielfältig Verwendung finden.

In Berlin werden die wichtigsten Personalien in den Organigrammen der Senats- und Bezirksverwaltungen aufgeführt und dezentral veröffentlicht. Diese wichtigen Informationen liegen in hoher Aktualität vor und können im Internet eingesehen werden. Organigramme werden in der Berliner Verwaltung bisher mit Microsoft Power-Point oder ähnlichen Programmen erstellt und im PDF-Format veröffentlicht. Der Aufbau und die Darstellung von Rollen und Beziehungen zwischen den einzelnen Organisationen folgen dabei keinen behördenübergreifenden, einheitlichen Standards. Organigramme werden bisher ausschließlich mit einem Fokus auf die Lesbarkeit für Menschen erstellt.

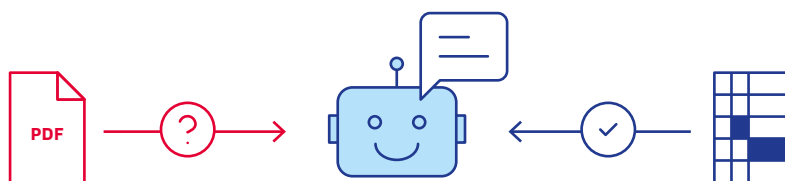


Abb. 3 Informationen aus einer PDF können von Computerprogrammen nur sehr begrenzt verarbeitet werden.

Öffnet man beispielsweise das Organigramm der Senatskanzlei, kann ein Mensch auf einen Blick recht schnell erfassen, wer den Posten *Regierender Bürgermeister von Berlin* innehat: Kai Wegner. Auch Franziska Giffey findet sich in diesem Organigramm, als *Bürgermeisterin von Berlin* (also als seine Stellvertreterin). Durch einen Computer, bzw. Programmcode, ist dieses PDF nur schwer automatisiert auslesbar. Würden die Informationen der Organigramme dagegen in einer strukturierten Tabelle vorliegen, hätte der Computer es schon deutlich

einfacher. Wir erinnern uns hier zur Verbildlichung an das Beispiel des Chatbots Bobbi aus der Einleitung, der fälschlicherweise Franziska Giffey als Bürgermeisterin benannt hat. Er lehrte uns bereits, dass für Digitalprojekte dieser Art die Verfügbarkeit von qualitativ hochwertigen, aktuellen und insbesondere auch maschinenlesbaren Informationen unabdingbar ist.

Allerdings haben auch strukturierte, maschinenlesbare Tabellen ihre Grenzen. Im Fall der Organigramme ist das recht offensichtlich: Die Verwaltungsstrukturen sind sehr komplex und geprägt von Hierarchien und Abhängigkeiten, die sich in einer tabellarischen Übersicht nur schwer abbilden lassen.

Ein weiteres Problem ist die Existenz vieler isolierter Datensätze in der Verwaltung. Ein Beispiel verdeutlicht dies: Angenommen, die Senatskanzlei veröffentlicht neben ihrem PDF-Organigramm eine maschinenlesbare Tabelle mit den Namen der Personen und deren Positionen. Ein darauf basierender Chatbot könnte auf die Frage nach Franziska Giffey's Position mit „Bürgermeisterin von Berlin“ (Stellvertreterin des Regierenden Bürgermeisters) antworten. Was der Bot jedoch nicht weiß: Sie hat auch den Posten als Senatorin für Wirtschaft, Energie und Betriebe (SenWEB) inne. Diese Information ist nicht explizit Teil des Organigramms der Senatskanzlei und somit dem Bot unbekannt.

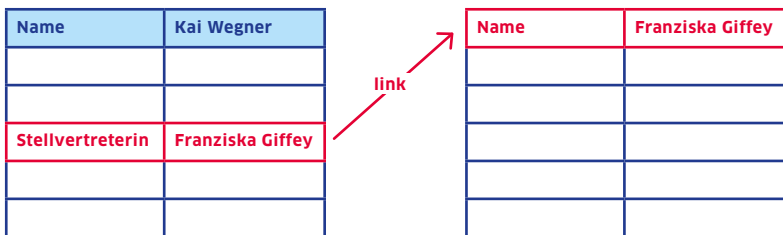


Abb. 4 Informationen, die eigentlich miteinander in Verbindung stehen sollten, liegen in unterschiedlichen Tabellen.

Um umfassend Auskunft geben zu können, müsste der Bot Zugang zu weiteren Datensätzen haben, in diesem Fall zum maschinenlesbaren Organigramm der SenWEB. Und in diesem müsste klar sein, dass mit Franziska Giffey die Bürgermeisterin und Senatorin Franziska Giffey gemeint ist und nicht zwei Personen mit zufällig gleichen Namen. Ein Chatbot, der auf eine Vielzahl unterschiedlicher, schwer vorauszusehender Fragen reagieren soll, stößt hier schnell an Grenzen.

3.2. Die Vision:

Das Wissensnetzwerk der Berliner Verwaltungshierarchie

Doch was wäre, wenn alle Informationen, die wir über bestimmte Sachverhalte, Objekte und Prozesse vorliegen haben, automatisch miteinander verknüpft werden könnten? Was wäre, wenn nach dem Regierenden Bürgermeister gefragt wird und der Bot nicht nur sagen kann, dass es sich dabei derzeit um Kai Wegner handelt? Er kann auch antworten, dass Wegner seit dem 26. April 2023 im Amt ist, dass sein Arbeitsort das Rote Rathaus ist, sein Geburtsort im Bezirk Spandau liegt, im Bezirk Spandau die CDU bei der Wiederholungswahl im Jahr 2023 stärkste Kraft war und vieles mehr.

Eine solche Verknüpfung von Daten ist möglich, wenn diese in einem speziellen Format bereitgestellt werden: Linked Open Data. Hinter diesem Begriff verbringt sich das Prinzip, dass Inhalte eines Datensatzes durch eindeutige Identifikatoren bezeichnet sind, die es erlauben, unmissverständlich und automatisiert Verbindungen herzustellen. So kann sich ein Netz an Informationen bilden, auch bezeichnet als Knowledge Graph, wie in Kapitel 2.2 bereits angerissen.

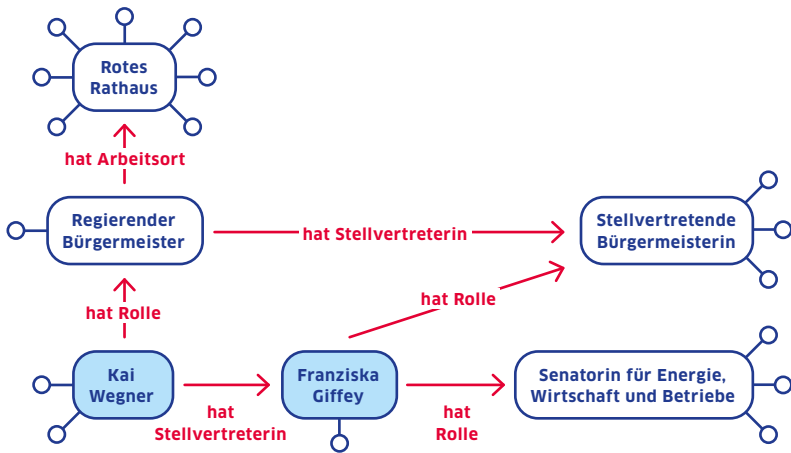


Abb. 5 Beispielhafte Skizze eines Knowledge Graph mit verlinkten Informationen zu den Rollen von Kai Wegner und Franziska Giffey

An diesem Punkt mag man sich fragen, weshalb all diese Anstrengungen unternommen werden, wenn fortschrittliche Sprachmodelle wie ChatGPT bereits über umfangreiches Wissen verfügen. Aber auch ChatGPT wurde mit Informationen trainiert, die wenige Zeit später schon wieder veraltet sind. Zudem basiert der Output von Sprachmodellen nicht auf logischen Schlussfolgerungen also „Wissen“, sondern auf der Wahrscheinlichkeit von dem zusammenhängenden Erscheinen von bestimmten Wörtern und Wortgruppen. Wie die in Kapitel 1.8 erwähnten Projekte aus der Schweiz zeigen, sind neuere Sprachmodelle aber durchaus in der Lage, aktuelle Information aus LOD auszulesen und in Kontext zu setzen. Darüber hinaus eröffnet ein gut strukturierter Wissensgraph eine Fülle neuer Fragestellungen und Erkenntnisse. Diese sind nicht nur für spezialisierte digitale Chat-Assistenten wie Bobbi relevant, um stets aktuell zu bleiben, sondern bilden eine wichtige Grundlage für verschiedenste digitale Anwendungen und Prozesse. Sie sind essenziell für die Digitalisierung der Verwaltung, die Beantwortung wissenschaftlicher Fragen oder das Training von Machine-Learning-Modellen.

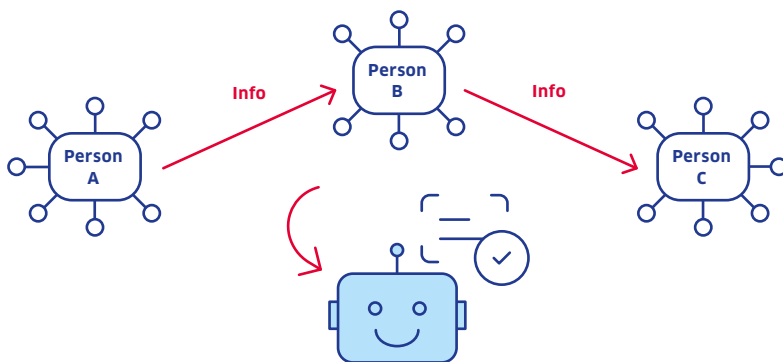


Abb. 6 Linked Open Data kann helfen, Informationen für Programme zugänglicher zu machen und Daten besser zu nutzen und zu interpretieren.

Inmitten der KI-Revolution wird also deutlich, dass wir für innovative Entwicklungen nicht nur maschinenlesbare Daten bereitstellen müssen. Es ist ebenso entscheidend, dass Informationen aktuell sind, sowie miteinander verknüpft und von Computern semantisch korrekt interpretiert werden können. Doch wie erstellen wir solche Daten?

3.3. Die Lösung: Das Organigramm-Tool als digitales Hilfsmittel

Im Sinne des Prototypings haben wir ein Tool entwickelt, mit welchem wir an diesem ganz konkreten Anwendungsfall – der Erstellung von maschinenlesbaren Organigrammen – das Konzept von LOD Schritt für Schritt erproben und erforschen können: Das Organigramm-Tool ist eine browserbasierte Webanwendung¹⁵ und ermöglicht über Standard-Textfelder Informationen zu Organisation und Personen einzugeben und in der bekannten Baumstruktur eines Organigramms

15 Das Tool ist im Internet frei einseh- und benutzbar: <https://organigramme.odis-berlin.de>. Der Source Code ist komplett Open Source und kann weiterverwendet werden: <https://github.com/technologiestiftung/organigramming-berlin>.

darzustellen. Die durch das Tool erstellten Organigramme können dann in Linked Open Data-fähigen Formaten (5 Sterne) exportiert werden, ohne viel technisches Hintergrundwissen zu benötigen.

Hier geht es zum Organigramm-Tool:

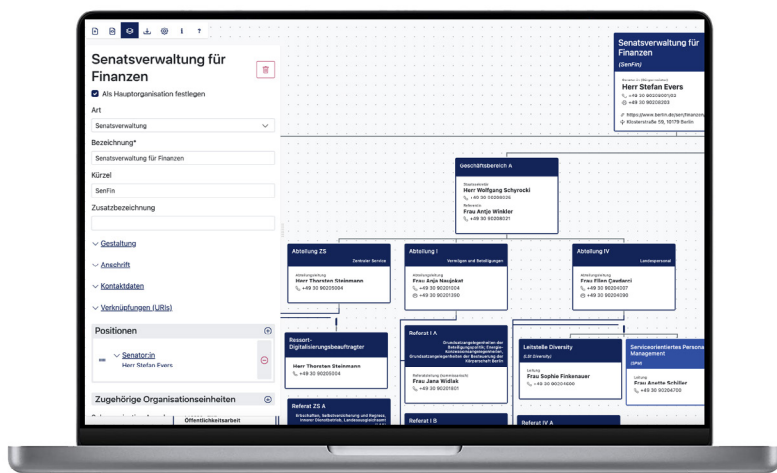


Abb. 7 Exemplarische Ansicht des prototypischen Organigramm-Tools

Gemeinsam mit den behördlichen Beauftragten für Open Data und Verwaltungsmitarbeitenden, die für die Erstellung von Organigrammen zuständig sind, fanden zudem Workshops und Arbeitstreffen statt. Gemeinsam wurden Standards, Positionen und Rollen in Berliner Verwaltungsorganigrammen erarbeitet, ohne die die Umsetzung

zu hochwertigen Linked Data nicht möglich ist. Im Ergebnis stehen zum Zeitpunkt der Veröffentlichung dieser Studie nicht nur ein funktionsfähiger Prototyp, sondern auch standardisierte Templates für die Organigramme der Senats- und Bezirksverwaltungen, eine damit zusammenhängende Ontologie für die Daten, sowie eine prototypische Datenbank zur Speicherung der Linked Data. Weiterhin konnten mit dem Tool in Zusammenarbeit mit der Verwaltung mehrere erste LOD-Organigramme angelegt und veröffentlicht werden.

Wie die Erstellung von LOD mit dem Organigramm-Tool im Rahmen der Fallstudie genau aussieht, und welche ersten Ableitungen sich daraus für die generelle Etablierung von LOD im Land Berlin ergeben, wird im folgenden Kapitel beschrieben.

Exkurs: Prototyping

Prototyping ist ein Prozess in der Entwicklung von Produkten, Dienstleistungen oder Software, um eine Idee greifbar zu machen und zu testen, bevor das endgültige Produkt entwickelt wird. Ein Prototyp kann von einer einfachen Skizze bis zu einer funktionalen Version reichen. Für uns als Open Data Informationsstelle Berlin (ODIS), sind Prototypen ein bewährtes Mittel, um zügig ins Machen zu kommen, neue Entwicklungen konkret in der Praxis zu erproben und am „lebenden Objekt“ zu lernen. Im Falle des Organigramm-Tools handelt es sich um eine sehr ausgereifte, funktionsfähige Web-Anwendung. Es wurde im Rahmen der Arbeitsgruppe der behördlichen Beauftragten für Open Data im Land Berlin und bei anderen Veranstaltungen regelmäßig im Entwicklungsprozess vorgestellt und basierend auf dem Feedback von potenziellen Nutzer:innen und von Linked-Data-Expert:innen Schritt für Schritt weiterentwickelt.

3.4. Linked Open Data erstellen: Vom PDF-Organigramm zum Triple Store

Das Datenqualitätsmodell von Tim Berners-Lee (siehe Kapitel 2.2), das von der einfachen Online-Verfügbarkeit von Textdokumenten (ein Stern) bis hin zur Einbindung in das Web der Daten, also Linked Data (fünf Sterne) reicht, bietet eine klare Richtschnur für die Bewertung und Verbesserung der Zugänglichkeit sowie der technischen Nutzbarkeit von Datensätzen.

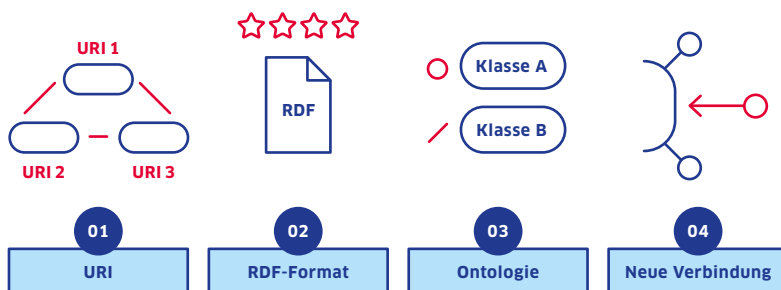


Abb. 8 Die vier LOD-Prinzipien auf einen Blick

Um die höchste Bewertung von fünf Sternen zu erreichen, müssen die Daten nicht nur maschinenlesbar sein, sondern auch den grundlegenden Prinzipien von LOD entsprechen:

1. Die Daten verwenden sogenannte Unique Resource Identifier, kurz **URIs**, zum Beschreiben von Objekten, Sachverhalten oder Personen und ihren Eigenschaften sowie Beziehungen zueinander.
2. Die Daten liegen im **RDF-Datenformat** vor, das Informationen in einer Art Satzform, inklusive Subjekt, Prädikat und Objekt abbildet.
3. Durch **Ontologien**, auch Vokabulare genannt, werden mithilfe von Klassen und Eigenschaften festgelegt, welche Beziehungen zwischen den Subjekten und Objekten möglich sind.

4. Durch die Verwendung der Prinzipien lassen sich **Verbindungen** zwischen Datensätzen herstellen. Dafür müssen diese im Netz erreichbar und mit einer speziellen Sprache abgefragt werden können.

Im Rahmen unseres Pilotprojekts konnten wir die Erstellung von 5-Sterne Linked Open Data am Beispiel der Berliner Organigramme durchspielen und alle vier Prinzipien umsetzen. Wie das in der Praxis aussieht, wird im Folgenden erklärt. Dafür widmen wir uns in den nächsten vier Kapiteln jeweils einem der Prinzipien.

3.4.1. Prinzip 1: Das Organigramm-Tool als URI-Generator und Editor

Eine zentrale Voraussetzung für Linked Open Data (LOD) ist, dass jede Ressource eindeutig identifizierbar ist. Dazu werden die schon erwähnten URIs (Uniform Resource Identifiers) verwendet, die es ermöglichen, Entitäten im Datensatz klar zu benennen und zu identifizieren.

Im Prinzip handelt sich dabei um beliebige, aber einmalige Zeichenfolgen. Dieses Prinzip ist zum Beispiel von ISBN-Nummern bekannt, die es ermöglichen, Bücher unverwechselbar zu identifizieren. Aber auch das World Wide Web basiert auf URIs: Jede Internetseite, also jede URL (Uniform Resource Locator), ist gleichzeitig eine URI. Die URL stellt eine eindeutige Verlinkung zu einer Website oder einer anderen über das Internet verfügbaren Datei dar. Wird sie in einen Internetbrowser eingegeben, kann auf die entsprechenden Dokumente zugegriffen werden. Im Fall von Linked Data sind URIs in erster Linie dazu da, die Entität eindeutig zu identifizieren. Meist handelt es sich bei den URIs in Linked Open Data aber gleichzeitig auch um URLs. Wenn eine solche URI im Webbrowser eingegeben wird, sollte sie zu einer nutzbaren Internet-Ressource führen, was die Transparenz, Interoperabilität und Nachnutzbarkeit der Daten gewährleistet.

Bei einem LOD-Organigramm bedeutet das, dass nicht nur das gesamte Organigramm eine einzige Webadresse bekommt, sondern jede einzelne Person, Rolle und Abteilung eine eigene, unverwechselbare URI erhält. Wichtig ist, dass diese Identifikatoren dauerhaft bestehen bleiben, unabhängig davon, ob sich die Rolle oder Position der Person im Laufe der Zeit verändert. Weiterhin sollten für gleiche Objekte und Personen, die in verschiedenen Datensätzen auftauchen, immer die gleichen URIs verwendet werden.

Ein Beispiel: Die Person Franziska Giffey bekommt folgende eindeutige URI zugewiesen.

<https://berlin.github.io/lod-organigram/person-ebfc6412ec>

Diese URI wird sowohl im maschinenlesbaren Organigramm der Senatsverwaltung für Wirtschaft, Energie und Betriebe verwendet, um dort ihre Position als Senatorin festzuhalten, als auch im Organigramm der Senatskanzlei, in der sie mit der Rolle der Bürgermeisterin verzeichnet ist. Die beiden Datensätze können so miteinander verlinkt werden. Für Maschinen ist anhand der URI eindeutig klar: Es handelt sich um ein und dieselbe Person.



Abb. 9 Anlegen einer URI im Organigramm-Tool für Franziska Giffey und Verknüpfung mit der existierenden Wikidata-URI.

Wie wurde dieses Prinzip in der Fallstudie mit dem Organigramm-Tool umgesetzt?

Wenn ein Organigramm im Tool angelegt wird, wird für jede Organisation, Position und Person eine URI automatisch generiert. Diese URI ist in der Regel eine zufällig erstellte Abfolge von Zahlen und Buchstaben. Die URIs können aber auch per Hand angelegt und bearbeitet werden. Weiterhin bietet das Tool die Möglichkeit, bereits existierende URIs von der bestehenden, offenen Datenbank Wikidata zu verknüpfen. Warum das relevant ist, greifen wir in Kapitel 3.5.2 auf.

3.4.2. Prinzip 2: Organigramm-Daten als Triples

Ein weiterer zentraler Aspekt für Linked Data ist die Organisation von Informationen bzw. der Daten in einer speziellen, dreiteiligen Struktur. Diese sogenannten Triples bestehen aus einem Subjekt, einem Prädikat und einem Objekt:

- Das Subjekt repräsentiert die Ressource oder Entität, auf die sich die Aussage bezieht. Dies kann zum Beispiel eine Person oder eine Abteilung sein, bezeichnet durch eine URI.
- Das Objekt repräsentiert einen Wert oder eine weitere Ressource, die mit dem Subjekt in Beziehung steht. Das könnte zum Beispiel ein Zahlenwert, wie die Höhe des jährlichen Etats sein, aber auch eine weitere Entität, wie zum Beispiel eine Person oder eine berufliche Position. Diese Entität wird dann ebenfalls durch eine URI identifiziert.
- Als Verbindung von Subjekt und Objekt gibt das Prädikat an, welche Art von Beziehung zwischen Subjekt und dem Objekt besteht und welche Informationen sich aus der Verbindung ergeben. Das Prädikat ist ebenfalls durch eine URI dargestellt; diese werden in einer sogenannten Ontologie, bzw. einem Vokabular festgehalten.

Eine Triple aus dem Datensatz eines LOD-Organigramms könnte zum Beispiel so aussehen:

<https://berlin.github.io/lod-organigram/organigram-20c5cdb22c>

(**Subjekt:** URI für die Organisationseinheit SenWEB)

<https://berlin.github.io/lod-vocabulary/berorgs#SenatorIn>

(**Prädikat:** URI für die Rolle “Senator:in”)

<https://berlin.github.io/lod-organigram/person-ebfc6412ec>

(**Objekt:** URI für die Person Franziska Giffey)

Aussage dieses Triples: Franziska Giffey ist Senatorin für Wirtschaft, Energie und Betriebe.

Ein Datensatz besteht für gewöhnlich aus sehr vielen Triples. Diese werden in einer standardisierten Form als Textdatei abgespeichert. Diese Form nennt sich RDF (Resource Description Framework). Es gibt verschiedene Dateiformate für Datensätze, die in Form von RDF abgespeichert werden sollen. Zwei besonders nutzer:innenfreundliche, weil auch gut menschenlesbare Formate, sind Turtle (Dateiendung .ttl) und JSON-LD (Dateiendung .jsonld).

Wie wurde dieses Prinzip in der Fallstudie mit dem Organigramm-Tool umgesetzt?

Ein in unserem Organigramm-Tool erstelltes Organigramm kann in RDF-Dateiformaten, zum Beispiel als Turtle-Datei, heruntergeladen werden. Alle Informationen des Organigramms sind in dieser Datei als Triples gespeichert.

```

@prefix schema: <https://schema.org/>.

organigram:person-ebfc6412ec a vcard:Individual;
    owl:sameAs <http://www.wikidata.org/entity/
    Q19279839>;
    vcard:family-name „Giffey“;
    vcard:given-name „Franziska“;
    vcard:tel „9013-8100“;
    org:holds organigram:position-c04bfa5fcc;
    schema:gender schema:Female.
organigram:position-c04bfa5fcc a org:Post;
    |   org:role berorgs:SenatorIn.

```

Abb. 10 Ausschnitt aus einer Turtle-Datei, die Informationen zu Franziska Giffey als Senatorin im RDF-Format, also als Dreisätze, enthält.

3.4.3. Prinzip 3: Eine Ontologie speziell für Berliner Organigramme

Eine weitere Voraussetzung für Linked Data ist die Verwendung von standardisierten Ontologien, also von spezialisierten Vokabularen zur Beschreibung der Daten. Sie bieten eine formalisierte Darstellung von Begriffen und deren Beziehungen in einem bestimmten Wissensbereich. Ontologien können als Schemata verstanden werden, die die Struktur und Bedeutung von Daten in einem bestimmten Kontext festlegen. Sie ermöglichen, dass Computerprogramme die Bedeutung von Daten „verstehen“.

Ontologien für allgemeine, weit verbreitete Datenkonzepte, die für viele Anwendungsbereiche relevant sind, werden von Standardisierungsorganisationen wie dem World Wide Web Consortium (W3C) oder Forschungsgruppen und Fachdatenbanken erarbeitet, festgelegt und

öffentlich dokumentiert. Diese standardisierten Ontologien dienen als Grundlage für den Austausch und die Verknüpfung von Daten im Web und sind bewusst breit gehalten, um in verschiedenen Domänen nutzbar zu sein. Ein Beispiel für eine existierende Ontologie, die häufig zur Beschreibung von Personen verwendet wird, heißt „vCard“. Ein Web-eintrag beschreibt, wie die Ontologie benutzt wird und bietet einen Überblick über ihre Begriffe, also Beziehungen und Informationen, die mit ihrem Vokabular abgebildet werden können¹⁶. Bei vCard sind das zum Beispiel die Namen, Anschrift und Kontaktmöglichkeiten der jeweiligen Person. Die Ontologie stellt somit auch URIs bereit, die benötigt werden, um Informationen als Triples auszudrücken.

Die Wiederverwendung bestehender Ontologien ist wichtig, um Interoperabilität zwischen Datensätzen zu gewährleisten, Redundanz zu vermeiden und eine bessere maschinelle Lesbarkeit durch etablierte Standards sicherzustellen. Für Fachthemen, die sich nicht ausreichend über bestehende Ontologien abdecken lassen, müssen neue eigene Ontologien erstellt werden.

Wie wurde dieses Prinzip in der Fallstudie mit dem Organigramm-Tool umgesetzt?

Diesem Prinzip folgend, bedient sich unser Organigramm-Tool sowohl bestehender Ontologien, als auch einer von uns speziell für Berlin verfassten. Wir nutzen unter anderem die weit verbreitete „vCard“-Ontologie, um Personen zu beschreiben und die „org“-Ontologie, um die Organisationsstrukturen darzustellen. Um die Beziehungen verschiedener Abteilungen und Mitarbeitenden der Berliner Verwaltung akkurat abbilden zu können, entwickelten wir für die Organigramme auch ein eigenes Vokabular mit dem Namen „berorgs“. Unser Web-eintrag für diese Ontologie in Form eines GitHub-Repositories definiert Berlin-spezifische Einrichtungen und Positionen und stellt die

16 Offizielle Dokumentation der vCard-Ontologie des W3C: <https://www.w3.org/TR/vcard-rdf/>

unmissverständlichen URIs für die Organigramme bereit¹⁷, zum Beispiel eine eindeutige URI für die Rolle *SenatorIn*:

<https://berlin.github.io/lod-vocabulary/berorgs/SenatorIn>.

3.4.4. Prinzip 4: Organigramme mit Daten im Web verlinken

Die eigentlichen Vorteile durch Linked Open Data, werden durch dieses letzte Prinzip eröffnet. Hierbei geht es darum, Verbindungen zwischen den Informationen verschiedener Datensätze herzustellen, also die Daten miteinander zu verlinken. Die Triples müssen dafür in einer Datenbank, dem sogenannten Triple Store gespeichert werden. Ein Triple Store ist eine Datenbank, die speziell darauf ausgelegt ist, RDF-Daten und die Beziehungen zwischen verschiedenen Datenpunkten zu verwalten. Anders als herkömmliche Datenbanken ist ein Triple Store darauf optimiert Beziehungsnetzwerke effizient zu speichern und zu durchsuchen.

Durch die Verbindung solcher Datenbanken entsteht ein globales Wissensnetzwerk, das es ermöglicht, neue Erkenntnisse aus verknüpften Datenquellen zu gewinnen. Dieses Netzwerk eröffnet völlig neue Möglichkeiten für Forschung, Analysen und die Erstellung innovativer Anwendungen.

Wie wurde dieses Prinzip in der Fallstudie mit dem Organigramm-Tool umgesetzt?

Wie bereits beschrieben, können die im Organigramm-Tool erstellten Organigramme in RDF-Dateiformaten heruntergeladen werden. Diese RDF-Dateien können nun in einem Triple Store abgelegt werden. Das Organigramm-Tool selbst, hat keine direkte Verbindung bzw. keinen direkt integrierten Triple Store. Das bedeutet, dass Organigramme,

17 Dokumentation der für die Berliner Organigramme entworfenen Ontologie „berorgs“:
<https://berlin.github.io/lod-vocabulary/berorgs/>

die im Tool erstellt werden, nicht automatisiert in einem Triple Store abgespeichert werden. Wir haben für diese Fallstudie einen Triple Store separat aufgesetzt und manuell die Berliner Organigramme, die in Zusammenarbeit mit Mitarbeitenden der Berliner Verwaltung erstellt wurden, reingeladen. Dieser Triple Store ist über eine Weboberfläche erreichbar. Über diese Oberfläche können die Informationen der Organigramme mittels der Programmiersprache SPARQL abgefragt und mit weiteren im Netz verfügbaren Linked Data verknüpft werden, beispielsweise mit Wikidata (siehe Kapitel 3.5.2).

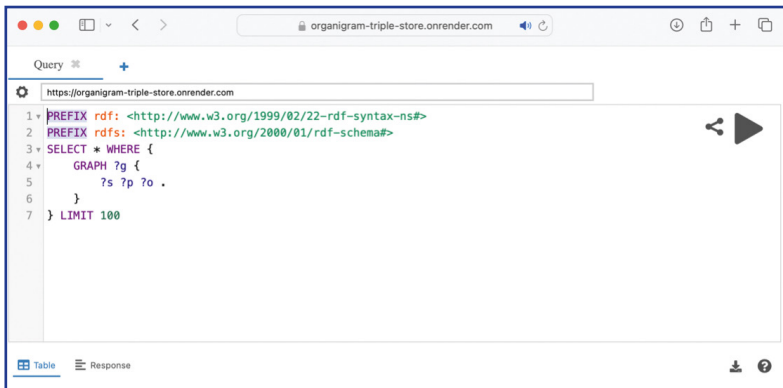


Abb. 11 Ansicht des SPARQL-Editors (der Weboberfläche) des Organigram-Triple Stores.

Während wir uns in diesem Kapitel dem Erstellen und Bereitstellen von Linked Open Data am Beispiel der Organigramme und mit Hilfe des eigens dafür erstellen Organigramm-Tools gewidmet haben, werden wir nun das Abfragen und somit die eigentliche Nutzung der Daten in den Fokus nehmen.

3.5. Linked Open Data nutzen: Die Organigramme befragen

Wie können die Linked Data der Organigramme verwendet und verarbeitet werden? In den folgenden beiden Unterkapiteln führen wir die Fallstudie fort und nutzen die neu gewonnene Datenbasis, um zum einen maschinell beispielhafte Fragen zur Struktur der Berliner Verwaltung mithilfe der Organigramme zu stellen und zum anderen, um die Informationen in den Organigrammen mit externen Informationsquellen zu verknüpfen und den Datenschatz somit weiter anzureichern.

Für diesen Teil der Fallstudie nutzen wir die LOD-Organigramme, die von den behördlichen Beauftragten für Open Data oder anderen Verwaltungsangestellten mit dem Organigramm-Tool erstellt und uns als RDF-Dateien für dieses Pilotprojekt im Jahr 2024 zur Verfügung gestellt wurden.¹⁸ Dabei handelt es sich um die Organigramme

- der Senatsverwaltungen für Finanzen (SenFin),
- der Senatsverwaltungen für Wissenschaft, Gesundheit, und Pflege (SenWGP),
- der Senatsverwaltung für Arbeit, Soziales, Gleichstellung, Integration Vielfalt und Antidiskriminierung (SenASGIVA),
- des Landesamts für Bürger- und Ordnungsangelegenheiten (LABO),
- des Landesamts für Einwanderung (LEA),
- des Bezirksamts Tempelhof-Schöneberg,
- des Bezirksamts Neukölln,
- des Bezirksamts Reinickendorf,
- des Bezirksamts Charlottenburg-Wilmersdorf und
- des Bezirksamts Steglitz-Zehlendorf.

18 Die Daten der Organigramme haben den Stand von Mitte Dezember 2024 und sind in der Detailtiefe Abbilder der zu diesem Zeitpunkt auf den Webseiten der jeweiligen Behörden veröffentlichten PDF-Organigramme. Die Organigramme der hier nicht aufgeführten Senatsverwaltungen und Bezirksamter, lagen uns bis zu diesem Zeitpunkt nicht maschinenlesbar vor.

Um die Informationen abzufragen, nutzen wir die schon erwähnte Abfragesprache SPARQL (Protocol and RDF Query Language). Das ist eine standardisierte, weit verbreitete Abfragesprache speziell den Umgang mit RDF-Daten bzw. Triples. Mit SPARQL können wir gezielte Fragen an die Organigramme in Programmiersprache übersetzen, z.B.: „Welche Personen sind in einer bestimmten Organisation tätig?“ oder „Welche Beziehungen existieren zwischen zwei Objekten?“.

Exkurs: SPARQL

Ähnlich wie SQL für relationale Datenbanken dient SPARQL dazu, Daten aus Triple Stores zu durchsuchen, zu filtern und zu verknüpfen. SPARQL-Abfragen folgen einer klaren Struktur:

Zunächst werden im **PREFIX** häufig verwendete URIs definiert, um die Abfrage übersichtlicher zu gestalten.

Danach legt die **SELECT**- oder **ASK**- Klausel fest, welche Daten abgefragt werden sollen. Während **SELECT** gezielte Datenpunkte zurückliefert, prüft **ASK**, ob bestimmte Daten vorhanden sind.

Im zentralen **WHERE**-Block werden Muster definiert, die auf das RDF-Datenmodell angewendet werden, um relevante Informationen zu filtern. Die **OPTIONAL**-Klausel ermöglicht es dabei, zusätzliche Informationen abzufragen, ohne dass deren Fehlen zum Ausschluss eines Ergebnisses führt.

Zusätzliche Klauseln wie **FILTER**, **ORDER BY** oder **LIMIT** ermöglichen eine gezielte Verfeinerung der Abfrage, indem sie beispielsweise Ergebnisse einschränken oder sortieren.

3.5.1. Anwendungsfall 1: Den Organigramm-Triple Store als Wissensspeicher über die Berliner Verwaltung nutzen

Die Organigramme können eine wertvolle Datenbasis sein, um Strukturen der Verwaltung zu analysieren. Ein zentraler Vorteil von LOD besteht darin, dass sich aus den RDF-Daten nicht nur einzelne Fakten extrahieren lassen, sondern auch komplexe Zusammenhänge sichtbar werden. So können beispielsweise Mehrfachfunktionen von Personen in verschiedenen Behörden identifiziert, organisatorische Abhängigkeiten analysiert oder Muster in der Verwaltungsstruktur erkannt werden.

Im Folgenden demonstrieren wir anhand konkreter SPARQL-Abfragen, wie der Triple Store als Wissensspeicher genutzt werden kann. Die gezeigten Beispiele dienen vorrangig Demonstrationszwecken und sollen eine Diskussion und Ideenfindung über die Potenziale der Datennutzung fördern.

Beispiel 1: Welche Schlüsselpositionen der Berliner Verwaltung sind unbesetzt?

Wo gibt es aktuell vakante Posten in der Leitungsebene der Berliner Verwaltung? Die folgende SPARQL-Abfrage filtert alle unbesetzten Posten heraus.

Die Kenntnis über unbesetzte Stellen könnte perspektivisch vielfältig genutzt werden. Die Entwicklung und Verteilung unbesetzter Stellen könnte in Zeiten des Fachkräftemangels eine interessante Information sowohl für potentielle Kandidat:innen, für Personaldienststellen als auch für die Stadtgesellschaft sein, um zu analysieren, ob die Verwaltung ausgeschriebene Stellen nicht besetzen kann oder ob bestimmte Stellen eine hohe Fluktuation aufweisen.

```

PREFIX org: <http://www.w3.org/ns/org#>
PREFIX vcard: <http://www.w3.org/2006/vcard/ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?position ?positionLabel ?orgLabel
WHERE {
    # Suche nach allen Positionen in den Organigrammen
    ?position a org:Post .
    OPTIONAL{?position rdfs:label ?positionLabel .}

    # Finde die jeweiligen Organisationseinheiten, zu den Positionen
    ?organization org:hasPost ?position .
    ?organization skos:prefLabel ?orgLabel .

    # Wähle nur Positionen aus, denen keine Person zugeordnet sind
    FILTER NOT EXISTS {
        ?person a vcard:Individual ;
            org:holds ?position .}
}

```

Output (vereinfacht)

	Position (URI)	Position (Label)	Organisationseinheit
1	position-ae6418601d	Amtsleitung	Wirtschaftsförderung
2	position-d5952a2924	Referatsleitung	Referat IV C (SenFIN)
3

Abb. 12 Diese SPARQL-Abfrage gibt alle unbesetzten Positionen der Organigramme aus. Als Indikator, dass eine Stelle nicht besetzt ist, wird hier geprüft, ob im Organigramm zu dieser Stelle eine Person angelegt wurde. Der Output-Teil zeigt einen vereinfacht dargestellten Ausschnitt der Ergebnisse, die auf die Abfrage ausgegeben werden.

Beispiel 2: Welche Beauftragtenrollen sind in welchen Behörden aktiv?

Welche Themen setzen die Verwaltungen politisch in den Fokus? Wie sind die Beauftragtenrollen angegliedert? Diese Fragen lassen sich mit einer SPARQL-Abfrage beantworten, die alle Beauftragten aus den Organigrammen ermittelt.

```
PREFIX org: <http://www.w3.org/ns/org#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?roleLabel ?parentOrgLabel
WHERE {
    # Rufe alle Rollen und übergeordneten Organisationseinheiten ab
    ?role org:hasPost ?position .
    ?role skos:prefLabel ?roleLabel .
    ?parentOrg org:hasSubOrganization ?role .
    ?parentOrg skos:prefLabel ?parentOrgLabel .

    # Nur nach Rollen filtern, die „Beauftrag“ im Namen enthalten
    FILTER (
        CONTAINS(LCASE(STR(?roleLabel)), "beauftrag")
    )
}
```

Output (vereinfacht)

	Rolle (Label)	Übergeordnete Einheit
1	Landesbeauftragte für Menschen mit Behinderung	Senatorin (SenASGIVA)
2	Ressort-Digitalisierungsbeauftragter	Abteilung ZS (SenFIN)
3

Abb.13 Diese SPARQL-Abfrage gibt alle Beauftragten-Rollen aus.

Diese Abfrage liefert eine Übersicht darüber, welche Beauftragten in welcher Behörde tätig sind. So lässt sich untersuchen, wo welche Themen vertreten werden – etwa durch Queer-Beauftragte, Seniorenbeauftragte oder Beauftragte für Antidiskriminierung. Dies erlaubt beispielsweise Einblicke in politische Prioritäten und Verwaltungsstrukturen der Stadt und gibt schnell einen Überblick über zuständige Personen.

3.5.2. Anwendungsfall 2: Organigramme mit externem Wissen verknüpfen

Die Organigramme der Berliner Verwaltung bieten bereits wertvolle Einblicke. Doch das volle Potenzial von Linked Open Data (LOD) entfaltet sich erst durch die Verknüpfung mit externen Wissensquellen. Indem wir die Organigramme mit zusätzlichen Daten – beispielsweise aus Wikidata – kombinieren, lassen sich neue Perspektiven gewinnen und die Analyse vertiefen.

Exkurs: Wikidata

Wikidata ist eine freie, offene Wissensdatenbank, die von der Wikimedia Foundation betrieben wird. Wikidata wurde entwickelt, um Wikipedia und andere Wikimedia-Projekte zu unterstützen, indem es eine zentrale Quelle für strukturierte Daten bietet. Der offizielle Start war im Oktober 2012. Heute dient es als eine zentrale, standardisierte Anlaufstelle für Daten innerhalb des Wikimedia-Ökosystems. Es gibt Millionen von Datenpunkten, die eine breite Palette von Themen abdecken, darunter Personen, Orte, Organisationen, wissenschaftliche Konzepte, historische Ereignisse und vieles mehr. Die Verknüpfung unserer Organigramm-Daten mit Wikidata erfolgt über eindeutige Identifikatoren, sogenannte Q-IDs, die Wikidata-Einträge repräsentieren. Durch SPARQL-Abfragen können wir gezielt externe Informationen aus Wikidata abrufen und mit unseren Daten kombinieren.

Externe Verknüpfungen ermöglichen es, Verwaltungsstrukturen nicht nur isoliert zu betrachten, sondern in einen größeren Kontext einzubetten. So lassen sich etwa Informationen über politische Entscheidungsträger, historische Entwicklungen von Behörden oder über regionale Vergleichsdaten in die Auswertung einbeziehen.

Beispielabfrage 3: Welche Informationen liegen zum politischen Background der Senator:innen vor?

Wie können wir die Transparenz und Zugänglichkeit der politischen Akteure in der Berliner Verwaltung verbessern? Die folgende SPARQL-Abfrage zeigt beispielhaft, wie man Informationen zu Senator:innen aus der Berliner Verwaltung abrufen kann. Diese Abfrage kombiniert interne Daten der Berliner Verwaltung mit Informationen aus Wikidata. So ist es möglich, Zugang zu detaillierteren, verknüpften Informationen zu bekommen, die in einem größeren Zusammenhang zur politischen Kommunikation und Verwaltungstransparenz stehen. Voraussetzung ist, dass für die Personen des Interesses ein Eintrag auf Wikidata vorliegt. Für Personen in hohen politischen Ämtern, wie den Senator:innen der Berliner Verwaltung, ist dies in der Regel der Fall.

Wir lassen uns hier ein auf Wikimedia hinterlegtes Bild der jeweiligen Person ausgeben. Weitere hinterlegte, interessante Informationen könnten zum Beispiel der politische Werdegang der Person, die Parteizugehörigkeit oder persönliche Informationen wie Alter und Geburtsort sein. Unsere Abfrage zeigt demonstrativ, wie wir eine Verbindung zu weiteren Informationen herstellen, die ihrerseits außerhalb von Wikidata zur Verfügung stehen. Abgefragt wird die Abgeordnetenwatch ID.¹⁹ Durch diese können unter anderem das Abstimmungsverhalten oder Antworten auf Bürger:innenfragen der Politiker:innen eingesehen werden.

19 Die Abgeordnetenwatch-ID ist eine eindeutige Kennung, die den Profilen auf der Plattform Abgeordnetenwatch.de zugeordnet wird. Abgeordnetenwatch.de ist eine deutsche Plattform, die es Bürger:innen ermöglicht, Fragen an ihre gewählten Vertreter:innen im Bundestag, in Landtagen sowie in Kommunalparlamenten zu stellen. Die Abgeordneten können diese Fragen beantworten, und ihre Antworten sind öffentlich zugänglich: <https://www.abgeordnetenwatch.de>.

```

PREFIX org: <http://www.w3.org/ns/org#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX berorgs: <https://berlin.github.io/lod-vocabulary/berorgs/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>

```

```

SELECT ?name ?wikidataID ?image ?abgeordnetenwatchID
# Nur nach Personen filtern, die die Rolle „SenatorIn“ haben
WHERE {
    ?person org:holds ?position ;
            rdfs:label ?name ;
            owl:sameAs ?wikidataID .
    ?position org:role berorgs:SenatorIn ;
    # Frage die Bild-URL und die ID anhand der Wikidata-Property ID ab
    SERVICE <https://query.wikidata.org/sparql> {
        ?wikidataID wdt:P18 ?image ;
                    wdt:P5355 ?abgeordnetenwatchID .
    }
}

```

Output (vereinfacht)



	Name	Abgeordnetenwatch ID	Wikimedia-Bild
1	Stefan Evers	28882	
2	Cansel Kiziltepe	79244	
3

Abb. 14 Diese SPARQL-Abfrage verbindet die Organigramm-Daten mit Daten von Wikidata. Abgefragt werden ein Bild und die Abgeordnetenwatch-ID aller Senator:innen.

3.5.3. Ausblick: Linked Data leichter nutzen mit KI

KI-Sprachmodelle könnten nicht nur von Linked Open Data als solide Datenbasis profitieren und Chatbots wie Bobbi schlauer machen. Sie können auch eine entscheidende Rolle bei der Verbesserung der Zugänglichkeit dieser Daten spielen. Anwendungen wie ChatGPT und die dahinterstehenden Modelle sind in der Lage, Code zu schreiben und zu interpretieren, wodurch sie natürliche Sprache in SPARQL-Abfragen umwandeln und die Ergebnisse entsprechend verarbeiten können. Dies erleichtert den Zugang zu Linked Open Data in Zukunft vermutlich erheblich, da Nutzer:innen nicht mehr direkt mit der komplexen Abfragesprache interagieren müssen. Auch die oben gezeigten SPARQL-Queries wurden mit einer KI-Unterstützung erstellt, die vorher mit den entsprechenden Ontologien „angelernt“ wurde. Es ist denkbar, spezielle angelernte KI-Agenten, die auf die Ontologien der Organigramme trainiert sind, auch für die breite Nutzung aufzusetzen und bereitzustellen. Da in der KI nur die Query erstellt wird und keine Daten direkt verarbeitet werden, kommt dieses Vorgehen sogar für sensible Daten in Frage.

Anschauliche Beispiele für die Erstellung von Abfragen mit KI sind das Projekt SPINACH²⁰ und das bereits erwähnte ZüriLinkedGPT (siehe Kapitel 2.4). Dabei wird transparent gemacht, wie die KI zur SPARQL-Abfrage und den erhaltenen Ergebnissen gelangt ist, sodass Nutzer:innen nachvollziehen können, wie das Ergebnis zustande kommt. Darüber hinaus könnten KI-Sprachmodelle dabei helfen, neu generierte Daten zu verarbeiten und für die schnelle Erstellung von Anwendungen oder Analysen nutzbar zu machen, zum Beispiel um Datenvisualisierungen zu erstellen.

20 SPARQL-basiertes Abfrage-Tool für Wikidata (SPINACH) der Stanford University:
<https://spinach.genie.stanford.edu>

4. Erkenntnisse und Handlungsempfehlungen

Mit dem Vorhaben zu Linked Open Data am Beispiel Berliner Organigramme wurde ein Stück Pionierarbeit für das Land Berlin geleistet. Genau wie Open Data an sich ist auch das Thema Linked Open Data nicht als Projekt zu verstehen, das für einen gewissen Zeitraum läuft und dann abgeschlossen wird, sondern als sich ständig entwickelndes, lernendes Vorhaben. Für die Erprobung von LOD können wir an dieser Stelle aber eine Zwischenbilanz ziehen. Nun gilt es auf den Ergebnissen und dem gewonnen Wissen aufzubauen, um LOD in der Verwaltung sukzessive zu entwickeln.

Dieses Pilotprojekt hat eine Reihe an Erkenntnissen gebracht, die für die nächsten Schritte, andere LOD-Projekte und strategische Entscheidungen seitens der Verwaltung relevant sind – auch über die Grenzen Berlins hinaus für Bund, Länder und Kommunen. Im Folgenden werden die gesammelten Erkenntnisse nicht nur erläutert, sondern auch mit konkreten Handlungsempfehlungen verknüpft.

4.1. Kompetenzen: Notwendiger Perspektivwechsel von Menschen- zu Maschinenzentriertheit

Eine Auseinandersetzung mit LOD benötigt viel Zeit für den Einstieg – sowohl um zu verstehen, wie Daten entsprechend der LOD-Prinzipien strukturiert werden können, als auch in Hinblick auf die Datennutzung zum Beispiel in Form der Abfragesprache SPARQL. Unser Projektteam, das anfangs mit Grundkenntnissen zu den Themen Linked Data und Semantic Web dieses Pilotprojekt initiiert hat, kann inzwischen einen stetigen Erkenntnisgewinn und eine damit hohe Lernkurve durch die Projektarbeit und intensive Beschäftigung mit dem Thema vorweisen. Im Umkehrschluss stellt sich die Frage, wie eine Vielzahl an Personen der Verwaltung entsprechende Kenntnisse im Bereich LOD aufbauen können bzw. welche Personen und Stellen

Kompetenzen benötigen, um eine Etablierung von LOD in der Berliner Verwaltung möglich zu machen. Eine sehr projektspezifische Herausforderung war die Vermittlung der Projektziele an die Nutzenden des Organigramm-Tools. In einigen Fällen wurde das Tool lediglich als ein Versuch verstanden, PDF-Organigramme einheitlich zu erstellen, nicht als Innovations- und Explorationsvorhaben für hochqualitative Daten. Dieses Missverständnis und geringere Gestaltungsmöglichkeiten gegenüber den bekannten Tools wie Visio oder PowerPoint führten zu weniger Bereitschaft, bei dem Projekt mitzuwirken.

Handlungsempfehlung

Ohne Kompetenzaufbau wird LOD sich nicht in der Verwaltung etablieren können. Verwaltungsmitarbeitende in den datenhaltenden Fachämtern können durch neu zu schaffende Schulungskonzepte an LOD herangeführt werden. Aus Sicht des Projektteams ist es allerdings noch essenzieller, gezielt auch Fachkräfte mit den erforderlichen Skills im Bereich Datenmanagement und Datenverarbeitung einzustellen. Diese fungieren als unterstützende Ressource für die Fachämter und helfen beim Abbau von Hemmschwellen bei der Einarbeitung in Linked Data. Ein gutes Beispiel hierfür ist die Senatsverwaltung für Finanzen, die Data Scientists eingestellt hat, die unter anderem bei der Veröffentlichung von LOD technisch unterstützen. Sobald solche spezialisierten Stellen in mehreren Behörden etabliert sind, sollten Austauschformate und Arbeitskreise geschaffen werden, um die Zusammenarbeit zu fördern und für Synergien und Standardisierung zu sorgen.

Datenverarbeitende Fachämter und Verwaltungen mit Datenbezug sollten trotzdem in die Verantwortung zum Kompetenzaufbau genommen werden. Ein Verständnis von LOD erfordert im wesentlichsten Schritt eine Auseinandersetzung mit der Frage, wie Maschinen bzw. Programme Informationen lesen und interpretieren und wie sehr sie sich von einer menschlichen Perspektive unterscheiden. Dieser Perspektivwechsel ist nicht nur wichtig für die Arbeit mit LOD, er ist auch essenziell für das Verständnis von Datenqualität. Dies gilt insbesondere, um angesichts der Entwicklungen im Bereich der Künstlichen Intelligenz

besser verstehen zu können, wie Informationen bereitgestellt werden müssen. Ein Fokus auf Datenkompetenz und Datenqualität im Allgemeinen sollte deswegen auch unter den Mitarbeitenden, die regelmäßig mit Daten arbeiten, stärker verankert werden. Nur so kann ein flächendeckendes Verständnis für die Bedeutung von Datenqualität in der Verwaltung aufgebaut und langfristig gefestigt werden.

4.2. Standards:

Organigramme als unerwartet komplexes Fallbeispiel

Verwaltungsorganigramme enthalten verschachtelte Informationen zu Zuständigkeiten und Hierarchien. Damit sind sie ein idealer Use-Case um die Vorteile gut strukturierter LOD hervorzuheben.

Gleichzeitig sind Organigramme aber „menschengemacht“ und historisch sowie politisch gewachsen. Die Bandbreite an unterschiedlichen Darstellungsformen und die Komplexität von Organigrammen waren und sind weiter sehr hoch und werden insbesondere im Versuch deutlich, Organigramme verschiedener Senatsverwaltungen oder Bezirke nebeneinander zu legen und zu vergleichen. Um sicherzustellen, dass Maschinen oder Software keine falschen Informationen erhalten, ist es jedoch unerlässlich, eine möglichst einheitliche Sprache bei der Erstellung der Organigramme zu verwenden. So haben wir uns mit nicht unerheblichem, und für uns vorher nicht richtig eingeschätztem zeitlichem Aufwand bemüht, gemeinsam mit den behördlichen Beauftragten für Open Data die Logiken von Organigrammen zu durchdringen, standardisierte Hierarchien zu entwickeln und Vergleichbarkeit herzustellen.

Neben einer Standardisierung der Organigramme war auch die Frage nach einer einheitlichen Sprache für die Beschreibung der Informationen zentral. Zu Beginn des Pilotprojekts stellte sich die Frage, auf welche Ontologie wir aufsetzen können. Da es bislang wenig Erfahrungen in der konkreten Umsetzung von LOD mit Verwaltungsdaten gibt, war es schwierig auf eine bestimmte bundesweite Ontologie aufzubauen. Zwar konnten wir auf erste kontrollierte Vokabulare wie Org, Schema

oder dterms zurückgreifen. In der Praxis mussten wir aber selbst ein Vokabular – beschränkt auf unseren Use-Case – entwickeln. Der Aufbau und die Pflege der Ontologie, auch aufgrund stetig neu hinzukommender Rollen, waren zeitintensiv und erfolgten aufgrund fehlender bestehender Referenzrahmen in einem wenig gefestigtem Umfeld. In der Berliner Verwaltung gibt es bisher keine Zuständigkeiten für die Entwicklung und Pflege von Ontologien. Dies erschwert die Erstellung berlinspezifischer Standards. Ohne eine zentrale Koordination besteht das Risiko von Redundanzen und inkonsistenten Strukturen.

Handlungsempfehlung

Für LOD-Projekte bedarf es einer eindeutigen Ontologie bzw. gemeinsamen Semantik, die eine konsistente Interpretation und Verknüpfung von Daten aus verschiedenen Quellen ermöglicht. Für Berlin wäre es sinnvoll, eine zentrale Zuständigkeit zu schaffen, die die Entwicklung von Ontologien für spezifische Anforderungen koordiniert und anstößt und gleichzeitig die Einhaltung internationaler Standards sicherstellt. Ein solcher Ansatz würde nicht nur die Interoperabilität der Daten fördern, sondern auch den Verwaltungsaufwand reduzieren und eine effizientere Zusammenarbeit ermöglichen. Mindestens aber muss es einen zentralen Ort geben, an dem bestehende, in Berlin genutzte Ontologien, aufgefunden werden. Wir empfehlen dafür das Anlegen eines oder mehrerer Repositorien, die Informationen über die Ontologien zum Beispiel als RDF-Dateien enthalten. Um die Ontologien besser auffindbar zu machen, könnten diese auch zentral im Berliner Open Data Portal eingetragen werden.

Auch wenn die Komplexität des Organigramm-Themas sich als sehr hoch herausgestellt hat, können wir Organigramme als ersten Use Case für LOD für andere interessierte Kommunen, Länder und dem Bund empfehlen. Dies nicht zuletzt, weil auf unsere Erfahrungen und entwickelte Ontologie aufgesetzt und das Organigramm-Tool nachgenutzt werden kann.

4.3. Infrastruktur: Lösungen über den Prototypenstatus hinaus benötigt

Eine technische Infrastruktur, insbesondere Triple Stores, sind eine Grundvoraussetzung für die Umsetzung von Linked Data, da sie die Verknüpfung und Abfrage semantisch strukturierter Daten ermöglichen. Aktuell besitzt die Verwaltung keinen offiziellen Triple Store, stattdessen wurde ein prototypischer Triple Store für dieses Projekt angelegt, der aber nicht für den dauerhaften Einsatz gedacht ist. In der Berliner Verwaltung stellt die Einrichtung eines Triple Stores eine Herausforderung dar, da sie Fragen nach Zuständigkeiten und Ressourcen aufwirft. Die Daten der Verwaltung liegen überwiegend dezentralisiert vor – Senatsverwaltungen und Bezirksämter verwalten ihre Informationen und betreiben IT-Infrastruktur unabhängig voneinander. Auch das Berliner Open-Data-Portal ist „lediglich“ ein Metadatenportal und zum jetzigen Zeitpunkt existieren keine zentralen Datenbanken oder Cloud-Speicher für die Datensätze. Die Frage lautet daher: Soll Berlin einen zentralen Triple Store bereitstellen, in dem alle Daten zusammengeführt werden, oder sollen die einzelnen Häuser jeweils ihre eigenen Triple Stores aufbauen?

Eine weitere Herausforderung betrifft die Handhabung der URIs. Zunächst muss geklärt werden, wer innerhalb der Verwaltung die Zuständigkeit für die Vergabe und Verwaltung von URIs hat. Ohne ein solches Vorgehen drohen Redundanzen, widersprüchliche Identifikatoren oder gar Kollisionen, bei denen eine URI für verschiedene Entitäten verwendet wird. Ein weiteres Problem ergibt sich aus der Dauerhaftigkeit der URIs. Es muss sichergestellt werden, dass die URIs aufgelöst und über lange Zeit hinweg stabil bleiben, selbst wenn sich organisatorische Strukturen oder Zuständigkeiten ändern. Dies erfordert klare Richtlinien für die Pflege und Archivierung von URIs sowie technische und organisatorische Maßnahmen, um diese Stabilität zu gewährleisten.

Handlungsempfehlung

Ein zentraler Ansatz erscheint uns sinnvoll, da es derzeit keine spezifischen Anforderungen der einzelnen Senats- oder Bezirksverwaltungen gibt, die eine dezentrale Speicherung und Einrichtung von Infrastruktursilos rechtfertigen würden. Der geplante Data Hub im Land Berlin sollte hier die benötigte technische Infrastruktur anbieten. Ein in dieser Datenplattform integrierter, zentraler Triple Store könnte die Interoperabilität verbessern, den Verwaltungsaufwand und Kosten reduzieren, und eine einheitliche Grundlage für die Umsetzung von Linked Open Data schaffen. Hierfür muss geklärt werden, wer einen offiziellen Triple Store für das Land Berlin betreiben und verwalten kann. Die Prozesse und Rahmenbedingungen zur Nutzung des Triple Stores durch Datenbereitstellende muss außerdem in einer zu erarbeitenden Berliner Data Governance bedacht werden.

Die Einführung einer klaren Governance ist auch deswegen zu empfehlen, damit geregelt wird, welche Einheit innerhalb der Berliner Verwaltung die Vergabe und Verwaltung von URIs koordiniert. Es sollte geprüft werden, ob ein zentrales URI-Register eingerichtet werden kann. Zudem wären automatisierte Tools hilfreich, analog zu unserem Ansatz im Organigramm-Tool, welche die URI-Generierung standardisieren und die Verknüpfung mit externen URIs, wie denen von Wikidata, ermöglichen.

4.4. Kritische Masse und Vernetzung: Wir brauchen mehr Linked Data

Damit Linked Open Data (LOD) in der Berliner Verwaltung seine volle Wirkung entfalten kann, ist es entscheidend, dass eine Vielzahl relevanter, verknüpfbaren Datensätze zur Verfügung steht. Aktuell sind bis auf die Organigramme dieses Pilotprojekts keine offiziellen Datensätze der Berliner Verwaltung als LOD verfügbar. Die Geometrien der lebensweltlich orientierten Räume (LOR) Berlins, geografische Unterheiten, die als Grundlage für Planung, Prognose und Beobachtung demografischer und sozialer Entwicklungen dienen, wurden von

BerlinOnline veröffentlicht und stehen als eine verknüpfbare Datenbasis bereit. Damit LOD kein Nischenthema bleibt, müssen verknüpfbare Daten aber systematisch verfügbar sein. Ohne eine kritische Menge an Linked Data, lässt sich die Kraft der Technologie nicht nutzen und das Momentum geht verloren. Die Umsetzung von LOD als ein Verwaltungsdigitalisierungsprojekt wird jedoch eine große Herausforderung für die meisten Stellen sein. Dabei kann es zu Beginn des Projekts passieren, dass noch gar nicht klar ist, mit welchen Akteuren und Akteurinnen eine Zusammenarbeit notwendig wird. Außerdem braucht LOD nach unseren Erfahrungen einen langen Atem. Kontinuität, die Notwendigkeit bereichsübergreifender Entscheidungen, sowie die notwendige Unterstützung durch die Leitungsebene. Es zeigt sich auch, dass LOD zumindest im deutschsprachigen Raum nach wie vor von einer überschaubaren interessierten Community vorangetrieben wird.

Handlungsempfehlung

In einem bereits initiierten Projekt, hat sich im Rahmen des Vierten Nationalen Aktionsplanes²¹ die Berliner Senatsverwaltung für Finanzen verpflichtet, den seit vielen Jahren als offenen Datensatz veröffentlichten Doppelhaushalt, als Linked Data im RDF-Format bereitzustellen. Dieses Projekt sollte unbedingt weiter unterstützt werden, denn es bietet nicht nur einen LOD-Datensatz für Berlin, sondern erarbeitet gleichzeitig wichtige Standards für Kommunal- und Bundesebene. Anschließend sollten Synergien zwischen den LOD-Organigrammen und den Haushaltsdaten hergestellt werden, z.B. durch einen Abgleich der URIs. Um die Herausforderung zu adressieren, ist es aber auch notwendig, weitere LOD-Projekte zu initiieren und Datensätze mit hohem Potenzial zur Veröffentlichung als LOD auszuwählen. Zu den priorisierten Datensätzen können insbesondere räumliche Bezüge, Infrastrukturdaten, Umweltdaten und Verkehrsdaten gehören. Diese

21 Der Nationale Aktionsplan (NAP) der Open Government Partnership (OGP) ist ein strategisches Dokument, in dem die Regierung konkrete Maßnahmen und Reformen zur Förderung von Transparenz, Bürgerbeteiligung festlegen: <https://www.open-government-deutschland.de/opengov-de/ogp/aktionsplaene-und-berichte/4-nap>

Datensätze sind nicht nur für spezifische Fachbereiche von Bedeutung, sondern auch für die Entwicklung übergreifender Analysen und Lösungen, etwa für die Verbesserung gesellschaftlich besonders relevanter Themen.

Ein wichtiger Baustein in der Umsetzung ist dabei eine grundlegende Kultur des langfristigen kollaborativen Arbeitens und die Bereitschaft, über den Tellerrand zu schauen. LOD wird sich nicht im stillen Kämmerlein entwickeln, sondern bedarf eines regelmäßigen Austauschs und Zusammenarbeitens verschiedener Akteure. Mit den behördlichen Beauftragten für Open Data im Land Berlin konnten wir einen kleinen Kreis an Linked-Data-Interessierten bilden. Für eine erfolgreiche Weiterentwicklung von LOD ist aber perspektivisch eine wachsende Community of Practice notwendig, die den Austausch pflegt und das Thema auf die digitalpolitische Agenda setzt. Um den Erfahrungsaustausch zu LOD in Deutschland zu stärken, haben wir gemeinsam mit der SenFin im Open Data Forum des Kompetenzzentrums Open Data (CCOD) ein Austauschformat²² initiiert.

22 Austausch für eine „Community of Practice“ im Open Data Forum von OpenCode: <https://discourse.opencode.de/t/community-of-practice-linked-open-data/3445/4>

5. Ausblick und Fazit

Das Projekt zur Einführung von Linked Open Data (LOD) in der Berliner Verwaltung ist noch nicht abgeschlossen. Es handelt sich um einen kontinuierlichen Prozess, der durch neue Erkenntnisse, Herausforderungen und technologische Entwicklungen ständig vorangetrieben wird. Unsere Erfahrungen und Erkenntnisse aus der Praxis sollen dabei dazu beitragen, die Grundlagen für eine langfristige Etablierung von LOD im Land Berlin zu schaffen. Hierzu wollen wir weiterhin einen kontinuierliche Austausch mit anderen Expert:innen und LOD-Enthusiast:innen pflegen, den wir durch die Teilnahme an Veranstaltungen wie der Piazza-Konferenz des Kompetenzzentrums Öffentliche IT (ÖFIT) und LOD-Barcamps bereits anregen konnten.

Der Weg zu einer vollständig vernetzten und datengetriebenen Verwaltung ist langfristig angelegt, und viele der implementierten Lösungen befinden sich noch in der Erprobung und Optimierung. Wir werden das Projekt der Organigramme weiterhin begleiten und die Umsetzung der Handlungsempfehlungen in der Berliner Verwaltung vorantreiben.

Ein wesentlicher Erfolgsfaktor für LOD wird darin liegen, konkrete und in der Verwaltungspraxis spürbare Mehrwerte zu schaffen. Die bereits bestehenden Organigrammdaten bieten hier eine wertvolle Grundlage. Aufbauend auf diesen Daten könnten innovative Use-Cases diskutiert werden, wie zum Beispiel eine Verknüpfung von Stellenbeschreibungen oder -profilen mit den Organigrammen. Die Daten könnten es auch ermöglichen, Fachverfahren zu optimieren, indem sie automatisiert zuständige Behörden, Abteilungen oder Ansprechpersonen für bestimmte Anliegen ermitteln. Beispielsweise könnten Eingangsanfragen in Bürgerämtern, Anträge auf Baugenehmigungen oder Fördermittel automatisch an die richtige Stelle weitergeleitet werden. Auch interne Verwaltungsprozesse wie Personalverwaltung, Ressourcenplanung oder Budgetsteuerung profitieren von aktuellen Organigrammdaten, da Zuständigkeiten klar hinterlegt sind. Ein weiterer vielversprechender Bereich ist der Ausbau digitaler Bürgerdienste und Verwaltungsportale. Viele digitale Verwaltungsangebote setzen voraus,

dass Bürger:innen oder Unternehmen wissen, welche Behörde oder Abteilung für ihr Anliegen zuständig ist. LOD-Organigramme könnten in Online-Services oder Chatbots integriert werden, um Nutzende schnell und automatisch zur richtigen Stelle zu leiten, anstatt sie durch komplexe Behördenstrukturen navigieren zu lassen.

Die geplante Veröffentlichung der Berliner Haushaltsdaten als LOD eröffnet zahlreiche Möglichkeiten, insbesondere die Verbindung von Haushaltsinformationen mit den bestehenden Organigrammen. So könnte beispielsweise analysiert werden, welche Haushaltsmittel von welchen Behörden oder Referaten verwaltet werden oder wie diese Mittel in den verschiedenen geografischen Einheiten von Berlin zum Einsatz kommen. Diese Art der Verknüpfung würde wertvolle Erkenntnisse für die Verwaltung sowie für die Zivilgesellschaft liefern und die Grundlage für eine bessere Steuerung und Kontrolle der öffentlichen Finanzen bieten.

Ein zukunftsweisender Bereich, der das Potenzial von LOD weiter steigern kann, ist die Integration von Künstlicher Intelligenz (KI). Wir möchten verstärkt untersuchen, wie KI und LOD zusammenarbeiten können, um die Interaktion mit und die Nutzung von LOD-Daten zu optimieren. Insbesondere stellen sich spannende Fragen wie: Wie können wir mit KI gezielt LOD-Daten abfragen? Inwieweit lassen sich Large Language Modelle dafür einsetzen, Daten zu pflegen oder komplexe Interaktionen zu ermöglichen? Wir sind gespannt, wie diese Technologien uns dabei unterstützen können, die Potenziale von Linked Open Data noch besser zu nutzen und neue innovative Ansätze für die Verwaltung zu entwickeln.

Am Ende dieser Fallstudie können wir zusammenfassen: Wir sehen das Potenzial, das in Linked Open Data (LOD) steckt und haben mit einigen begeisterten Menschen gesprochen und zusammengearbeitet, die das ebenso sehen. LOD bietet eine enorme Chance, Verwaltungsdaten zu neu zu nutzen und aus den bisherigen Informationssilos ein dynamisches Wissensnetzwerk zu schaffen.

Jedoch sehen wir auch die Herausforderungen, die mit der Umsetzung von LOD verbunden sind. Es ist ein komplexer Prozess, und wir sind uns bewusst, dass es nicht von heute auf morgen eine Vielzahl an LOD-Daten geben wird. Vielmehr müssen wir die Entwicklungen im Bereich LOD weiterhin aufmerksam verfolgen, besonders im Zusammenspiel mit neuen Technologien wie KI, und kontinuierlich mit konkreten Pilotprojekten und Experimenten lernen, wie wir das volle Potenzial ausschöpfen können. Es bedarf eines schrittweisen Aufbaus von Technik, Kompetenz und Rahmenbedingungen, um LOD in der Verwaltung langfristig zu etablieren.

Abbildungsverzeichnis

Abb. 1	Das Datenreifegrad-Modell nach Tim Berners-Lee.	18
Abb. 2	Ausschnitt aus dem PDF-Organigramm der Senatskanzlei (Stand: März 2024)	22
Abb. 3	Informationen aus einer PDF können von Computerprogrammen nur sehr begrenzt verarbeitet werden.	23
Abb. 4	Informationen, die eigentlich miteinander in Verbindung stehen sollten, liegen in unterschiedlichen Tabellen.	24
Abb. 5	Beispielhafte Skizze eines Knowledge Graph mit verlinkten Informationen zu den Rollen von Kai Wegner und Franziska Giffey	26
Abb. 6	Linked Open Data kann helfen, Informationen für Programme wie Bobbi zugänglicher zu machen und Daten besser zu nutzen und zu interpretieren.	27
Abb. 7	Exemplarische Ansicht des prototypischen Organigramm-Tools	28
Abb. 8	Die vier LOD-Prinzipien auf einen Blick	30
Abb. 9	Anlegen einer URI im Organigramm-Tool für Franziska Giffey und Verknüpfung mit der existierenden Wikidata-URI.	32
Abb. 10	Ausschnitt aus einer Turtle-Datei, die Informationen zu Franziska Giffey als Senatorin im RDF-Format, also als Dreisätze, enthält.	35
Abb. 11	Ansicht des SPARQL-Editors (der Weboberfläche) des Organigramm-Triple Stores.	38

- Abb. 12** Diese SPARQL-Abfrage gibt alle unbesetzten Positionen der Organigramme aus. Als Indikator, dass eine Stelle nicht besetzt ist, wird hier geprüft, ob im Organigramm zu dieser Stelle eine Person angelegt wurde. Der Output-Teil zeigt einen vereinfacht dargestellten Ausschnitt der Ergebnisse, die auf die Abfrage ausgegeben werden. 42
- Abb. 13** Diese SPARQL-Abfrage gibt alle Beauftragten-Rollen aus. 43
- Abb. 14** Diese SPARQL-Abfrage verbindet die Organigramm-Daten mit Daten von Wikidata. Abgefragt werden ein Bild und die Abgeordnetenwatch-ID aller Senator:innen. 46
- Abb. 15** Schematische Darstellung eines Knowledge Graphs: Mit einer Vielzahl an verknüpften Daten entstehen leistungsstarke Wissensnetzwerke 66

Glossar wichtiger Begriffe

HTTP

Hypertext Transfer Protocol. Grundlegendes Protokoll für die Kommunikation im World Wide Web. Es definiert die Regeln für den Austausch von Daten wie HTML-Dokumenten, Bildern, Videos und anderen Ressourcen zwischen einem Client (z. B. einem Webbrowser) und einem Server.

JSON

Java Script Object Notation. Ein leichtgewichtiges, textbasiertes Datenformat, das hauptsächlich für den Datenaustausch zwischen Systemen verwendet wird. JSON ist einfach zu lesen und zu schreiben, sowohl für Menschen als auch für Maschinen, und basiert auf einer Untergruppe der JavaScript-Programmiersprache.

Knowledge Graph / Wissensgraph

Eine strukturierte Darstellung von Wissen, die Informationen in Form von Entitäten (z. B. Personen, Orte, Dinge) und deren Beziehungen (z. B. „arbeitet für“, „ist Teil von“) organisiert. Diese Darstellung ermöglicht es, Daten in einer semantischen, maschinenlesbaren Weise zu speichern, zu verknüpfen und zu nutzen.

Ontologie

Set von Begriffen (Konzepte, Objekte) und Beziehungen (Verknüpfungen zwischen diesen Begriffen), das Wissen über eine Domäne beschreibt und eine gemeinsame Grundlage für Kommunikation und Datenverarbeitung bietet.

RDF

Resource Description Framework. Ein Standard der W3C (World Wide Web Consortium), der eine Methode zur Beschreibung von Daten und deren Beziehungen im Web bietet. RDF wird häufig verwendet, um Informationen im sogenannten Semantic Web strukturiert, vernetzt und maschinenlesbar darzustellen.

SPARQL

SPARQL steht für SPARQL Protocol and RDF Query Language. Eine standardisierte Abfragesprache des W3C (World Wide Web Consortium), die verwendet wird, um Daten aus RDF-Datenbanken (Resource Description Framework) abzufragen, zu manipulieren und zu analysieren.

Triple

Einfache Struktur, bestehend aus drei Elementen:

Subjekt: Das Ding, über das eine Aussage gemacht wird;

Prädikat: Die Eigenschaft oder Beziehung;

Objekt: Der Wert oder die Entität, mit der das Subjekt in Beziehung steht.

Triple-Store

Eine spezielle Art von Datenbank, die für das Speichern, Verwalten und Abfragen von Daten in Form von Tripeln optimiert ist.

URI

Uniform Resource Identifier. Standardisierte Methode zur eindeutigen Identifikation von Ressourcen im Internet oder in einem Informationssystem. Ein URI ist eine Zeichenkette, die verwendet wird, um eine Ressource (z. B. ein Dokument, eine Datei, eine Datenbank, ein Service oder eine Entität) zu benennen und gegebenenfalls auf sie zuzugreifen.

Repository

Ein Speicherort für Daten, Code oder Dokumente, der eine strukturierte Verwaltung und den Zugriff auf diese Ressourcen ermöglicht. In der Software- und Datenwelt kann ein Repository auch eine Plattform zur Versionskontrolle (z. B. Git-Repository) sein.

LOD**(Linked Open Data)**

Verknüpfte, öffentlich zugängliche Daten, die den Prinzipien von Linked Data folgen. Diese Daten sind maschinenlesbar und fördern die Interoperabilität zwischen verschiedenen Systemen und Anwendungen.

Semantic Web

Erweiterung des klassischen Webs, in dem Daten durch semantische Technologien verknüpft und maschinenlesbar gemacht werden. Ziel ist es, dass Maschinen Inhalte nicht nur darstellen, sondern auch verstehen und verarbeiten können.

W3C**(World Wide Web Consortium)**

Standardisierungsgremium für Webtechnologien, einschließlich Linked Data und Semantic Web. Es entwickelt und pflegt offene Standards, um die langfristige Entwicklung des Webs zu sichern.

Turtle**(Terse RDF Triple Language)**

Kompakte Syntax zur Darstellung von RDF-Tripeln. Sie erleichtert das Schreiben und Lesen von RDF-Daten im Vergleich zur XML-basierten RDF-Darstellung.

JSON-LD**(JSON for Linked Data)**

Erweiterung von JSON, die es ermöglicht, Daten mit semantischer Bedeutung darzustellen. Dadurch können Web-APIs und andere JSON-basierte Systeme einfach mit Linked Data-Technologien integriert werden.

SHACL**(Shapes Constraint Language)**

Standard zur Validierung von RDF-Datenstrukturen. Damit lassen sich Regeln definieren, um sicherzustellen, dass Daten einer bestimmten Struktur oder einem bestimmten Schema entsprechen.

OWL**(Web Ontology Language)**

Erweiterung von RDF zur Definition komplexer Ontologien und logischer Regeln. OWL ermöglicht es, formale Beziehungen zwischen Konzepten zu definieren und automatisierte Schlussfolgerungen zu ziehen.

Weiterführende Literatur und Leseempfehlungen

Europäische Kommission (2022): Linking Data. What does it mean?,
Offizielle Webseite: <https://data.europa.eu/en/publications/datastories/linking-data-what-does-it-mean>

Forschungsdaten.Info (2024): FAIRe Daten: <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/faire-daten>

Kanton Zürich: Linked Open Data auf dem Prüfstand,
<https://www.zh.ch/de/news-uebersicht/mitteilungen/2024/politik-staat/statistik/linked-open-data-auf-dem-pruefstand.html>

Open Data Informationsstelle Berlin (2024): ODIS-Blogpost:
Vom Informationssilo zum Wissensnetzwerk, <https://odis-berlin.de/aktuelles/2024-03-28-linked-open-data>

Open Data Informationsstelle Berlin (2024): ODIS-Blogpost
Verständliche Organigramme für alle, <https://odis-berlin.de/aktuelles/2024-05-021-organigramme-workshop>

Open Data Informationsstelle Berlin (2024): ODIS-Prototyp: Organigramm-Tool, <https://odis-berlin.de/projekte/2023-07-organigramm-tool>

Open Knowledge Foundation (2024): Open Data Ranking Deutschland
<https://opendataranking.de>

Senatskanzlei Berlin (2023): Berliner Open Data Strategie 2023
(Seite 10 ff.), <https://www.berlin.de/moderne-verwaltung/e-government/open-data/strategieprozess/artikel.1257333.php>

Sporny, Manu(2012): Youtube-Video: What is Linked Data?,
https://www.youtube.com/watch?v=4x_xzT5eF5Q

Willems, Christina (2024): Piazza Konferenz 2024 Rückblick
<https://piazza-konferenz.de/2024/12/17/linked-open-data-vernetzte-verwaltung>

Vom LOD zum Knowledge Graph

Mit vielen Datenbereitsteller:innen gemeinsam ans Ziel



Ein Subjekt in einem Triple kann in einem anderen Triple als Objekt fungieren, wenn es eine Beziehung zu einem Subjekt beschreibt.

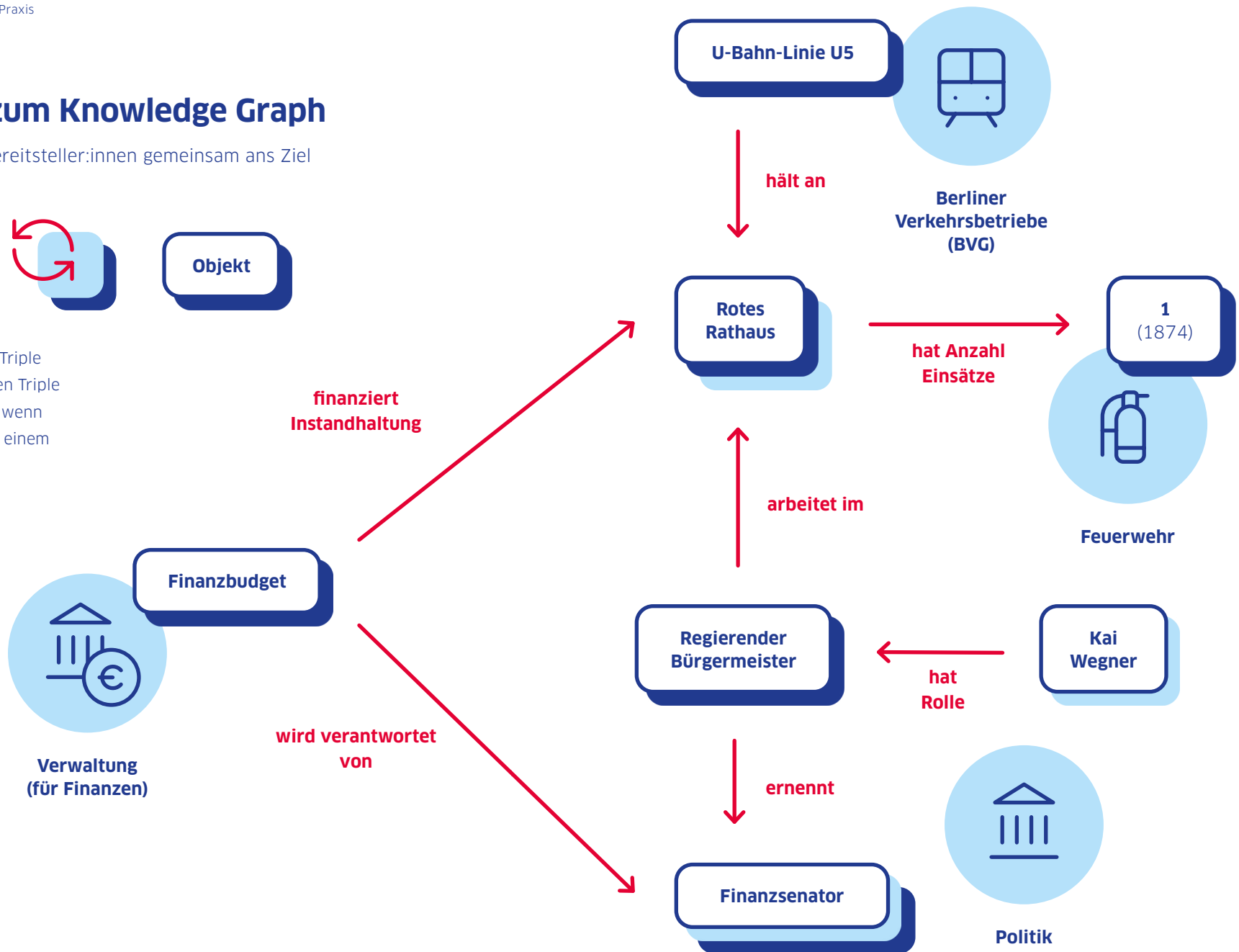


Abb. 15 Schematische Darstellung eines Knowledge Graphs: Mit einer Vielzahl an verknüpften Daten entstehen leistungsstarke Wissensnetzwerke.

Impressum

© Technologiestiftung Berlin

März 2025

Technologiestiftung Berlin

Grunewaldstraße 61-62

10825 Berlin

T +49 30 209 69 99 0

info@technologiestiftung-berlin.de

technologiestiftung-berlin.de

Open Data

Informationsstelle Berlin

odis@ts.berlin

odis-berlin.de

Autor:innen

Lisa Stubert

Klemens Maget

Max Eckert

Hans Hack

Infografiken

Anna Meide

Gestaltung

Anja Bender

Lektorat

Gesa Trojan

Pauline Boos

Die ODIS ist ein Projekt der
Technologiestiftung Berlin.

Gefördert von der Senatskanzlei

Berlin und der Investitionsbank

Berlin aus den Mitteln des Landes
Berlin.

This work is licensed under
a Creative Commons

Attribution-ShareAlike 4.0

International License

[https://creativecommons.org/
licenses/by-sa/4.0](https://creativecommons.org/licenses/by-sa/4.0)



**TECHNOLOGIE
STIFTUNG
BERLIN**

