

Use Case

Haushaltsdaten als

Linked Data

- Phase 1: Status Quo feststellen, vernetzen, Arbeitspakete schnüren
- Phase 2: Austausch, Vokabular, Veröffentlichung
- Phase 3: Community of Practice

Phase 1: Status Quo feststellen, vernetzen, Arbeitspakete schnüren

Als [Ergebnis](#) eines [Beteiligungsworkshops](#) zur Open-Data-Strategie des Landes Berlin stand die Forderung nach mehr Linked Data im Raum - der Berliner Finanz Senat nahm den Ball auf und setzte sich zum Ziel, die bereits [existierende Visualisierung](#) der Haushaltsdaten auch im Hintergrund "schön" zu machen -> im Linked Open Data Format.

<https://haushaltsdaten.odis-berlin.de/share>

In der ersten Phase des Projekts ging es vor allem darum zu schauen, was es schon an Initiativen innerhalb der Verwaltung gab, bei denen Informationen aus Fachverfahren direkt als Linked Data veröffentlicht werden. Zur Bestandsaufnahme gehörte auch der deutschlandweite Blick auf die Arbeit mit Haushaltsdaten und die Identifizierung möglicher Kooperationspartner:innen.

1. Bestandsaufnahme - alte und aktuelle Projekte

- Die Erfahrungen mit dem [Projekt offener Haushalt](#) haben gezeigt, dass es eine initiale Bereitschaft von Projektpartnern braucht, um nachhaltige Strukturen dafür innerhalb der Verwaltung aufzubauen - das heißt, eine zentrale Ansprechperson und technische Begleitung, damit die Erfahrungen und das Wissen innerhalb der Verwaltung bleiben und für Folgeprojekte und interessierte Menschen aus anderen Abteilungen genutzt werden kann.
- Auf internationaler Ebene gab es mit dem "[Linked Spending](#)" als Teil des [Open Spending Projekts](#) bereits 2014 einen Ansatz, samt Vokabular, um Haushalte europaweit verlinkbar zu machen. Nach dieses anfänglichen Hochtagen von Linked Data passierte erst einmal nichts - die automatisch umgewandelten Datensätze können aber als Skizze für das Vorhaben dienen. Zudem wurde ein [Paper](#) veröffentlicht, das die Umwandlung von Haushaltsdaten ein Datacube behandelt.
- In Thüringen forschen [Marianne Mauch](#) von der Uni Jena und [Felicita Löffler](#) vom Finanzministerium Thüringen an Wissensgraphen für die Verwaltung. Der Projektstand lässt sich im Kurzformat [hier](#), für einen konkreten Fall [hier](#), bzw. mit etwas mehr Kontext eingebettet im [Workshop von der Piazza-Konferenz](#) nachvollziehen.
- Die Senatskanzlei Schleswig-Holstein hat in Person von Jesper Zedlitz in den letzten Jahren bereits an der [Veröffentlichung von Daten als Linked Open Data](#) gearbeitet. Sein [Interesse an Haushaltsdaten](#) führte auch dazu, dass das Land Schleswig Holstein als Projektpartner gemeinsam mit der Senatsverwaltung für Finanzen in Berlin daran arbeitet, die Haushaltsdaten der beiden Länder verlinkbar zu machen und an einer Ontologie zu arbeiten.

- Die ODIS in Berlin ist seit 2023 damit beschäftigt, an einer Lösung zu arbeiten, wie man Organigramme in ein maschinenlesbares und gleichzeitig verlinkbares Format umwandeln kann. Die eigene Auseinandersetzung mit dem Thema Linked Data und die dazugehörige steile Lernkurve hat das Team in einer [tollen Präsentation](#) festgehalten, die sehr niedrigschwellig und anschaulich Linked Data erklärt und unbedingt empfohlen ist. Die von der ODIS schon erstellte [Ontologie](#) für die Berlin-spezifischen Einrichtungen und Positionen, kann als Grundlage für
- Knud Möller von [BerlinOpenData](#) hat bereits mit der Veröffentlichung von Linked Data über Github experimentiert. Seine Überlegungen gibt es als [Präsentation hier](#) - ein Blick in das Repository kann [hier](#) geworfen werden.

2. Barcamp als Auftakt

Um die Leute, die sich mit dem Thema beschäftigen, zu vernetzen und einen ersten Überblick über den aktuellen Status Quo zu ermitteln, haben wir gemeinsam mit dem Berliner Finanzsenat ein Barcamp ausgerichtet, zu dem Menschen aus der Wissenschaft, Zivilgesellschaft und auch direkt aus der Berliner Haushaltsabteilung eingeladen waren. Die Ergebnisse und Protokolle der einzelnen Beiträge können unserem [Etherpad](#) entnommen werden.

3. Teilnahme am 4. Nationalen Aktionsplan OGP

Gemeinsam mit dem Land Schleswig-Holstein wurde eine Teilnahme am 4. NAP OGP vereinbart. Der aktuelle Stand auf der Projektseite kann [hier](#) eingesehen werden. Im NAP heißt es:

Die Berliner Haushaltsdaten werden künftig als Linked Open Data (LOD) auf dem Open-Data-Portal des Landes zur Verfügung stehen. Ziel ist eine Verknüpfung, die Vollständigkeit und eine verbesserte Qualität der Daten. Gemeinsam mit dem Land Schleswig-Holstein soll zudem der Erfahrungsaustausch, die Verknüpfung und Skalierung des Vorhabens in die Wege geleitet werden.

4. "Ins Machen kommen"

Der aktuelle Stand des Projekts wurde bei der AG Open Data des Landes Berlin im März vorgestellt. Die Präsentation dazu findet sich hier: [AGOpenDataLOD.pdf](#). In einem nächsten Schritt geht es um das Mapping der Daten, bei dem wiederum ein Austausch mit Expert:innen durchgeführt wird.

Phase 2: Austausch, Vokabular, Veröffentlichung

1. Das passende Format für Haushaltsdaten

Zentral für die zweite Phase war eine Bestandsaufnahme mit den beteiligten Parteien. Dabei ging es konkret in ein moderiertes Gespräch mit Expert:innen aus der Haushaltsabteilung und den Menschen, die mit der technischen Umsetzung befasst sind.

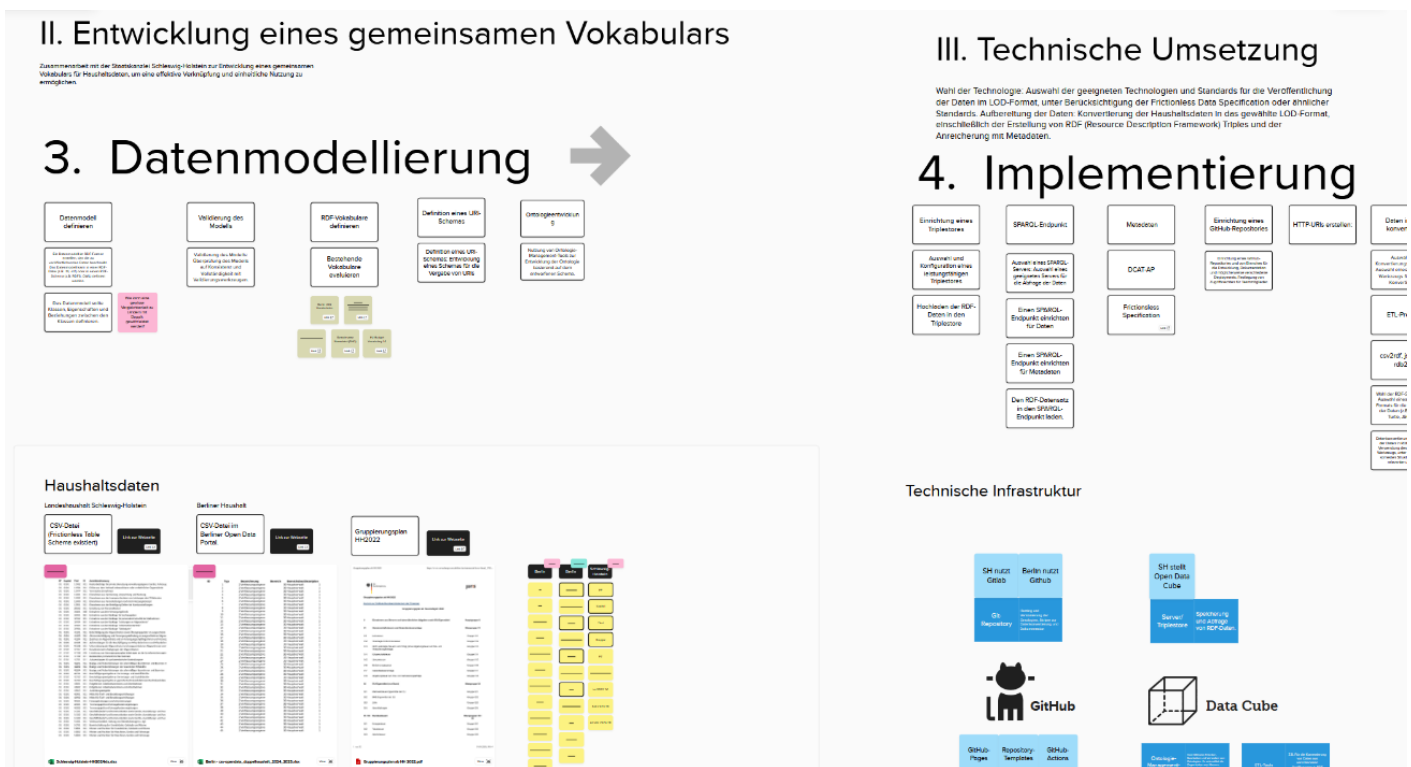


Bild: Ausschnitt aus dem gemeinsamen Arbeitsplan

Bei dem Gespräch wurden Unterschiede und Gemeinsamkeiten zwischen den Haushaltsdatensätzen der zwei Länder identifiziert und ein in Schleswig-Holstein entwickeltes Python-Skript für die Umwandlung von einer CSV zur RDF dem Team in Berlin zur Verfügung gestellt. Mit einigen Anpassungen konnten so die Haushaltsdaten beider Bundesländer in ein RDF-Format gegossen werden. Das Prinzip dahinter erklärt Jesper Zedlitz von der Staatskanzlei Schleswig-Holstein in einem [Blogbeitrag](#). Für die Haushaltsdaten wurde sich an der existierenden Vorarbeit von [Openspending](#) - [Linkedsending](#) orientiert, die bereits vor mehr als zehn Jahren Haushaltsdaten in [RDF Data Cubes](#) modelliert haben.

2. Verlinkung ermöglichen - gemeinsames Vokabular für Haushaltsdaten entwickeln

Im Rahmen des gemeinsamen Gesprächs wurde festgestellt, dass für die Daten aus den Landeshaushalten bislang kein passendes Vokabular, das die eigentliche Verlinkung ermöglicht, existiert und deshalb selbst auf Basis etwa der

Ontologie des EU Haushalts und den Beispielen von Linkedspending erstellt werden muss. Bei der Datenmodellierung geht es um die Definition von Datenpunkten und ihrer Bedeutungen. In einem Vokabular wird mithilfe von Klassen und Eigenschaften festgelegt, welche Beziehungen zwischen diesen Datenpunkten möglich sind. Ziel ist es, ein konsistentes Schema für die Identifizierung von Daten zu entwickeln, das die Eindeutigkeit sicherstellt und gleichzeitig flexibel genug ist, um regionale und lokale Besonderheiten zu berücksichtigen.

original	überarbeitet	original
Berlin	Berlin	Schleswig Holstein
ID	Beispiel	EP
Typ	Beispielbezeichnung	Kapitel
Bezeichnung	Kapitel	Titel
Bereich	Kapitelbezeichnung	Gruppe
Bereichsbezeichnung	Gruppe	FZ
Ergebnis	Gruppenbezeichnung	Zweckbezeichnung
Ergebnisbezeichnung	Titel	Ist 2022 T€

Bild: Beispiel für den Vergleich verschiedener Bereiche der Haushaltsdaten

Durch den Austausch zwischen den Haushaltsabteilungen der beiden Länder wurde festgelegt, worin der gemeinsame Teil zwischen den Haushaltsdatensätzen liegt oder wo gleiche Posten schlicht anders benannt sind. Für allgemeine Informationen sollte auf bereits etablierte Standards für Felder wie Jahreszahl etc. oder den Standard SDMX für statische Daten zurückgegriffen werden und so ein allgemeines Vokabular für Haushaltsdaten entstehen, das jeweils um landesspezifische Teile erweiterte werden kann.

3. Von RDF zu URIs

Die Definition von URIs bilden den Kern von Linked Open Data. Mittels einer URI lassen sich Daten online eindeutig identifizieren und wiederverwenden. Bei der Diskussion über die Verwendung von URIs zur Identifikation von Datensätzen zeigte sich der Ansatz, dass jede Verwaltungseinheit ihre eigenen Namensräume verwenden sollte, um Klarheit und Ordnung in der Datenorganisation zu gewährleisten. Der Namensraum gibt dabei Hinweis auf die Herkunft der Datenpunkte/des Vokabulars. Für die Haushaltsdaten hieß das konkret, dass die landesspezifischen Daten auf den Namespace des jeweiligen Bundeslands verweisen, während für die übergreifenden Daten ein einheitlicher "Bundes"-Namespace genutzt wird. Mangels einer existierenden Basis aus dem Bund, wurde für das allgemeine Haushaltsvokabular die Plattform der Open Knowledge Foundation Deutschland genutzt.

Ein Weg, um aus existierenden RDFs diese Identifier zu kreieren und über Github zugänglich zu machen, wurde von BerlinOnline entwickelt. Mit **Jinja-RDF** können leicht statische Seiten erzeugt und angepasst werden. Da Berlin derzeit nicht über einen Triple Store verfügt, wurden die **Haushaltsdaten selbst ebenfalls bei Github** veröffentlicht.

LOD-Haushalt

Datensatz

Typ: [void:Dataset](#)

Beschreibung: Budgetary data for Berlin as RDF / Linked Open Data.

zuerst erstellt am: 2024-10-17

zuletzt geändert am: 2025-04-07

Veröffentlichende Stelle: [Senatsverwaltung für Finanzen](#)

Homepage: <https://github.com/berlin/lod-budget>

Beispielresource: [hhbe:6d3eeee8-2066-506e-a73a-4d23c0f759a2](#) (hh:Titel)

Beispielresource: [hhbe:e1c72463-dd22-5f4e-9119-89f43c83af74](#) (hh:Einnahmetitel)

Beispielresource: [hhbe:047106de-d059-54c7-b8ae-a2b132d4f1c8](#) (cube:Observation)

Beispielresource: [hhbe:2d61d0af-86a4-5458-8af5-8946d94e75dd](#) (hh:Ausgabebetitel)

Beispielresource: [hhbe:dd19576c-00df-53f6-9721-ab8c0dc0227a](#) (cube:Observation)

Beispielresource: [hhbe:979e4bb7-0518-5a87-9693-d01f44c29099](#) (cube:Observation)

Genutztes Vokabular: [sdmxdim:](#)

Genutztes Vokabular: [qudt:](#)

Phase 3: Community of Practice