

Checkliste: Linked Open Data für Verwaltungen

PHASE 0

Linked Data Grundsätze verstehen

Ein Grundverständnis von Wissensgraphen und Linked Data ist bei der Projektorganisation hilfreich. Hier sind einige Materialien aufgeführt, die Konzepte veranschaulichen und Praxisbeispiele zeigen.

Materialien der ODIS: Blogbeitrag „Vom Informationssilo zum Wissensnetzwerk“, Fallstudie „LOD in der Praxis“.

Kostenloser Grundlagenkurs [Knowledge Graphs](#) beim Hasso Plattner Institut
Use-Case Linked Data aus der Stadt Zürich – Workshop, [Anwendungsbeispiel](#), [Erfahrungsbericht](#).

Selbstlerneinheit zu [Metadaten](#), [Semantic Web & SPARQL](#) vom QUADRIGA Datenkompetenzzentrum.

PHASE 1

Status quo erfassen – Projektumfang umreißen

Zunächst geht es darum festzulegen, welche Daten veröffentlicht werden sollen, die zuständigen Personen zu identifizieren und interne Unterstützungspotentiale auszuloten.

Welche Informationen sollen als Linked Data zur Verfügung gestellt werden, was ist die datenliefernde Stelle? Besteht die Möglichkeit die Daten direkt automatisiert aus einer Schnittstelle zu bekommen? In welchem Format und mit welchen Metadaten liegen die Daten vor? Gibt es interne Anwendungsszenarios, die an einer Aufbereitung bzw. besseren Verfügbarkeit der Daten Interesse haben könnten?

Gibt es Personen mit datenspezifischer Expertise in der Abteilung, die eingebunden bzw. für eine initiale Einschätzung gewonnen werden könnten? Wer ist die verantwortliche Person für das Fachverfahren/den Prozess aus dem die Daten entstehen? Gibt es Ansprechpersonen zum Thema Daten/Open Data in der Behörde selbst oder an einer übergeordneten Stelle, die man konsultieren könnte?

Mitstreiter:innen suchen

Nach der ersten internen Bestandsaufnahme kann der Kreis erweitert und recherchiert werden, ob ein ähnliches Projekt vielleicht bereits anderswo durchgeführt wurde oder wird.

Gab/gibt es Projekte mit ähnlichem Ansatz schon bei weiteren Behörden? Ein erster Überblick kann mithilfe von simpler Suchmaschinen- und Literaturrecherche geschafft werden. Weitere Ansatzpunkte sind Hinweise auf existierende Vokabulare für das Themenfeld etwa auf [europäischer](#) Ebene oder bei anderen [Anbietern](#). [Schema.org](#) und [wikidata](#) sind weitere Quellen für Schemata für strukturierte Daten. Möglicherweise gibt es bereits eine Ontologie, auf die für eine Veröffentlichung aufgebaut werden kann.

Linked Open Data folgt einem genuin kollaborativen Ansatz – je mehr weitere Stellen Daten verlinkbar bereitstellen, desto höher das Potential der Nutzung. Darum sollte geschaut werden, ob es vielleicht für dasselbe Vorhaben Menschen aus anderen Behörden/ Bundesländern gibt, mit denen man direkt gemeinsam an der Datenveröffentlichung zu demselben Thema arbeiten könnte, bzw. ob es in der Vergangenheit schon einmal ähnliche Ansätze gegeben hat. Erste Anlaufstellen sind u.a. Personen aus dem Daten-Ökosystem ([lokale Community](#), [Open Knowledge Foundation](#)) oder behördenübergreifende Arbeitsgruppen. Darüber hinaus gibt es Plattformen wie das [Open Data Forum](#) des CCOD oder das [regelmäßige Open Data Netzwerktreffen der Bertelsmannstiftung](#).

Projektplan erstellen und Arbeitspakete bilden

Gerade bei einem Thema wie Linked Data hilft es, sich zum Projektstart den möglichen Verlauf vor Augen zu führen und in Arbeitspakete mit Zuständigkeiten aufzuteilen.

Zunächst empfiehlt es sich, einen groben Ablaufplan von Auftaktmeeting bis hin zur Veröffentlichung zu erstellen:

1. Wie und in welchem Format werden die Daten bereitgestellt? Wer nutzt die Daten aktuell und könnte Feedback geben und perspektivisch von Linked Data profitieren?
2. Existiert ein Vokabular oder ein anderer Standard auf dem man aufbauen kann?
3. Wer kann ein Datenmodell erstellen – gibt es interne Kapazitäten, um institutionelles Wissen in Schemata zu gießen? Kann inhaltliche Expertise mit Programmierexpertise zusammengebracht werden, oder muss dieser Schritt beauftragt werden?
4. Wer kann zu dem Datenschema Feedback geben – andere Menschen mit inhaltlicher/technischer Expertise (siehe Ansprechpersonen)?
5. Wie sollen die Daten/ein mögliches Vokabular veröffentlicht werden – existiert eine Infrastruktur dafür auf Behörden-/Landesebene?

Modellierung und Datenarbeit

Bei der Modellierung gilt es, Fach- mit Programmierlogik zusammenzubringen – Visualisierungstools und ein klarer Ablaufplan für alle Beteiligten können die Arbeit erleichtern.

Gibt es kein direkt passendes Vokabular, liegt die Hauptarbeit in der Modellierung, also der Definition von Datenpunkten und ihrer Bedeutungen. Ziel ist es, ein konsistentes Schema für die Identifizierung von Daten zu entwickeln, das die Eindeutigkeit sicherstellt und gleichzeitig flexibel genug ist, um etwa regionale und lokale Besonderheiten zu berücksichtigen. Dabei ist es wichtig, frühzeitig zu klären, welcher **Namespace** genutzt werden soll, um eine eindeutige Identifikation zu gewährleisten.

Da hierfür sowohl Fachwissen als auch technische Kenntnisse erforderlich sind, wäre es sinnvoll, wenn dieser Schritt von IT-Kräften der Behörde selbst übernommen werden könnte. So kann Wissen für zukünftige Anwendungen, Anpassungen und Kooperationen aufgebaut werden.

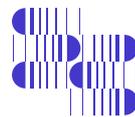
Feedback, Veröffentlichung und Outreach

Linked Data lebt vom Austausch – dieser sollte auch rund um die Veröffentlichung gesucht werden.

Vor der Veröffentlichung sollte das Schema in Feedbackschleifen gemeinsam mit Expert:innen betrachtet und angepasst werden. Hier ist Pragmatismus gefragt: Vokabulare lassen sich anpassen und erweitern. Der Fokus sollte daher darauf liegen, das Ergebnis zu veröffentlichen und zur Diskussion zu stellen, statt sich in Details zu verlieren.

Existiert eine Infrastruktur (Triple Store) für die Veröffentlichung und unmittelbare Abfrage von Linked Data? Hier können die Betreiber:innen der jeweils zuständigen Datenportale vermutlich Aufschluss geben. Ein alternativer Weg ist die **Veröffentlichung** über [Github](#) oder [OpenCode](#). Ansonsten können die Daten auch einfach in Linked Data Formaten (z.B. Turtle oder RDF) als Download angeboten werden.

Als finalen Schritt gilt es mögliche Anwender:innen, die am besten schon von Anfang an miteinbezogen wurden, anzusprechen und weitere interne und externe Datensätze zu identifizieren, die sich für eine Verlinkung anbieten. Zudem ist es hilfreich, eine Reihe an Beispiel-Queries, also Abfragen an die vorhandenen Daten, mitzuliefern.



Gemeinsam mit der Staatskanzlei *Schleswig-Holstein* hat die Senatsverwaltung für Finanzen in Berlin ihre Haushaltsdaten als *Linked Open Data* veröffentlicht und ein *Vokabular* entwickelt. Hier sind fünf Gelingensbedingungen, die wir aus dem Projekt mitgenommen haben.

1. Interne und externe Sparringpartner suchen

Ein zentraler Baustein für den Projektfortschritt war die Möglichkeit, sich bei Fragen niedrigschwellig auszutauschen. Dabei konnte die Berliner Behörde intern auf die Unterstützung der Berliner Open-Data-Informationsstelle und des Betreibers des Berliner Datenportals zählen sowie das Thema in der Berliner AG Open Data diskutieren. Extern gab es mit der Open Knowledge Foundation eine Anlaufstelle, die am Ball blieb und die Fortschritte begleitete. Über den Nationalen Aktionsplan *Open Government Partnership* war zudem die Staatskanzlei Schleswig-Holstein involviert. Sie konnte vor allem bei Entwicklungsfragen weiterhelfen und stellte einen ersten Entwurf für das Haushaltsdatenvokabular bereit.

2. Behörden- mit Programmierwissen zusammenbringen

Bei der Modellierung geht es darum, die Verwaltungslogik in ein möglichst allgemeingültiges und anknüpfbares Datenschema zu übersetzen. Hilfreich war hierbei, dass es für Haushaltsdaten bereits bundeseinheitliche Standards gibt, auf die man sich als Diskussionsgrundlage berufen konnte. Wichtig war außerdem ein Grundverständnis der jeweiligen Gegenüber auf der Haushalts- sowie auf der Programmierseite. Trotzdem bleibt es ein Vorantasten, das auf Behördenwissen und viel Kommunikation angewiesen ist.

3. Kompetenzen inhouse aufbauen

In der Senatsverwaltung für Finanzen gibt es zwei Data Scientists, von denen eine Person am LOD-Projekt mitarbeiten konnte. Dadurch musste die Programmierleistung nicht extra ausgeschrieben werden und die Kommunikation lief über den kurzen Dienstweg. Zudem konnte so Wissen in der Behörde aufgebaut werden, das etwa für Weiterentwicklungen, interne Fortbildungen und Nachfragen aus anderen Senatsverwaltungen genutzt werden kann – ein entscheidender Punkt, damit es nicht bei einem Leuchtturmprojekt bleibt.

4. Direkt in den Austausch gehen

Zum Auftakt fand ein Barcamp statt, bei dem sich die Projektteams mit verschiedenen Akteur:innen aus Deutschland persönlich treffen und ergebnisoffen zu Linked Data austauschen konnten. Aus unterschiedlichen Blickwinkeln auf Haushaltsdaten zu blicken und von anderen Erfahrungen mit Linked Data zu hören, war ein guter Start, um mit der Arbeit loszulegen. Bei verschiedenen Konferenzen und Austauschformaten wurde zudem über den Projektfortschritt gesprochen und so Feedback eingeholt, anstatt allein im stillen Kämmerlein zu arbeiten.

5. Aufmerksamkeit auf die Infrastruktur richten

Die Arbeit an der Datengrundlage ist kein Thema, mit dem sich auf der Entscheider:innenebene unmittelbar Lorbeeren ernten lassen. Ergebnis und Wirkung sind nicht so leicht zu visualisieren wie ein blinkendes Dashboard, ein Chatbot oder das Durchschneiden eines Eröffnungsbandes. Umso wichtiger ist es, positives Feedback in die Führungsebene zu tragen und an öffentlichkeitswirksamen Fachkonferenzen teilzunehmen oder eine Selbstverpflichtung über Programme wie der *OGP* einzugehen. Auf das große Potential von Wissensgraphen und abgestimmten Vokabularen für die Verwaltung weist auch die Projektgruppe „Semantische Interoperabilität“ des IT-Planungsrats in ihrem *Abschlussbericht* hin.