



Kanton Zürich
Direktion der Justiz und des Innern
Statistisches Amt

Schöntalstrasse 5
8090 Zürich
Paketadresse:
Schöntalstrasse 5
8004 Zürich
Telefon 043 259 75 00
Telefax 043 259 75 69
zh.ch/statistisches-amt

Linked Open Data, Abschlussbericht der Erprobungsphase

30.08.2024



Einleitung

Der Kanton Zürich hat die Publikation von Linked Data während zwei Jahren konzipiert und erprobt. Dieser Bericht enthält eine Zusammenfassung der Ergebnisse und Erkenntnisse. Der Bericht setzt Kenntnisse zur Funktionsweise von Linked Data voraus. Wer sich dazu erst einlesen möchte, kann das auf der [Website von LINDAS](#) tun.

Ausgangslage und Ziel

Linked Open Data (LOD) gilt als beste Form, offene Daten bereitzustellen (Tim Berners-Lee, vgl. 5stardata.info/de). Dem Kanton Zürich fehlt bisher das Know-How und die Infrastruktur, um LOD zu publizieren. Daher besteht der Bestand an Open Government Data des Kantons Zürich bislang aus tabellarischen Daten, mehrheitlich im Format csv. Der Kanton Zürich publiziert daher «3-Star-Data».

Besser wäre gemäss diesem Modell Daten in RDF-Format, ein semantischer Standard, welcher neben der Daten auch die Metadaten bereits enthält (4-Star-Data). Und Linked Open Data, welches die Verknüpfung der Datensätze untereinander bereits enthält und weitere Verknüpfung jederzeit zulässt (5-Star-Data). Erklärend könnte man sagen: 3-Star-Data ermöglicht die maschinelle Verarbeitung von Daten, 4-Star-Data kontextualisiert die Daten mittels eingebetteter Metadaten und 5-Star-Data lässt die vernetzte Auswertung dieser Daten unmittelbar zu.

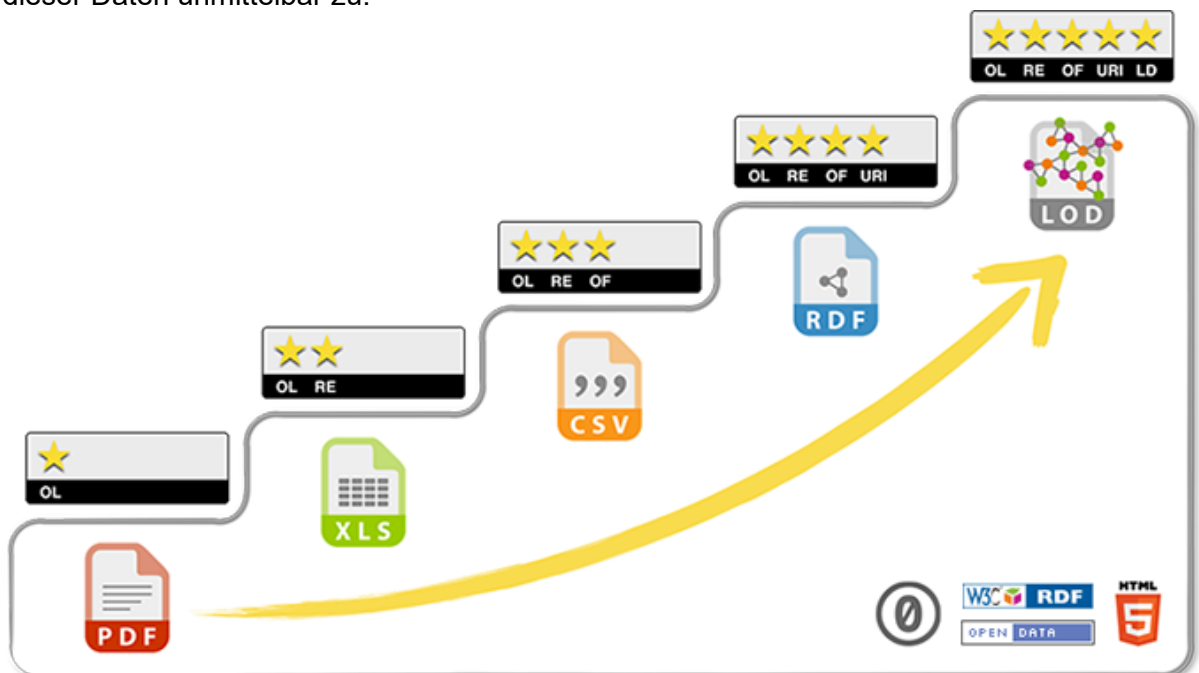


Abbildung 1: 5-Sterne-Modell für Open Data, Tim Berners-Lee (5stardata.info, 08.11.2023)



In einem ersten Versuch wollte eine Arbeitsgruppe aus interessierten Datenspezialist*innen Erfahrungswerte in der Aufbereitung und der Publikation von LOD sammeln. Ziel war es, Antworten zu finden auf die Fragen:

1. Wie aufwändig ist es, bestehende Daten in Linked Data aufzubereiten?
2. Welche Infrastruktur ist nötig, um Linked Data zu publizieren?
3. Wird Linked Open Data von der Open Data Community genutzt?

Aus diesen Erfahrungswerten wollen der Fachausschuss und die Fach- und Koordinationsstelle für Open Government Data ihre Strategie bezüglich LOD ableiten.

Organisation

Für die Erprobung von LOD hat sich eine AG Linked Open Data gebildet. Die AG LOD traf sich im Zeitraum März 2022 bis Oktober 2023. Die Projektadministration übernahm unterstützend die Fach- und Koordinationsstelle OGD, die Unterlagen und erarbeiteten Grundlagen sind auf einem [Github-Repository der Fach- und Koordinationsstelle OGD](#) abgelegt, jedoch nicht öffentlich zugänglich.

Leistungen und Ergebnisse

Vernetzung

Zur Gründung der AG LOD kam es über eine Anfrage bei der OGD-Fachstelle des Staatsarchivs. Das Staatsarchiv suchte nach eine Publikationslösung für die in RDF konvertierten [Ehedaten](#). Daraus ergab sich die Idee einer Infoveranstaltung zum Thema LOD, um das diesbezügliche Bedürfnis in der Kantonsverwaltung abzuholen. An der Infoveranstaltung präsentierte Hansjörg Stalder von der Statistik Stadt Zürich die Erfahrungen, welche in der Stadt Zürich mit Linked Open Data gemacht wurden. Die Stadt verwendet LOD zur Veröffentlichung sämtlicher statistischer Fakten und nutzt dafür aktuell ebenfalls die Bundesinfrastruktur LINDAS. Aus interessierten Teilnehmenden bildete sich nach der Infoveranstaltung die Arbeitsgruppe, in welcher auch mehrere Personen von Statistik Stadt Zürich vertreten waren.

Zu erwähnen ist noch, dass bereits im Vorfeld beim Kanton erste Erfahrungen mit LOD gesammelt werden konnten. In einem Projekt des Staatsarchivs und des Statistischen Amtes zusammen mit Markus Pilzecker, wurde versucht, die Kantonsratsmitglieder seit 1803 als Linked Data aufzubereiten.



Infrastruktur

Die Arbeitsgruppe hat die Nutzung von LINDAS, dem Linked Data Service des Bundes für öffentliche Verwaltungen der Schweiz, während 1.5 Jahren (Juni 2022 bis Dezember 2023) getestet. Neben LINDAS gibt es verschiedene weitere Möglichkeiten, LOD zu publizieren.

LINDAS strebt an, alle LOD der öffentlichen Verwaltungen in der Schweiz zu publizieren. Der Service will damit einen zentralen Zugang zu LOD der Schweiz bieten. Teil des Angebots ist die Software Cube Creator, mit welchem Daten als Linked Data aufbereitet werden können. LINDAS übernimmt das Hosting der LOD und bietet Datennutzenden [verschiedene Zugänge auf die LOD](#).

Der Service ist kostenpflichtig, das Pricing richtet sich in drei Stufen nach der Anzahl Triples. Skaleneffekte bei wachsendem Angebot an LOD werden nicht weitergegeben, der Service bleibt deshalb auf absehbare Zeit sehr teuer. Die Erprobung hat die AG LOD mit dem kleinsten Angebot umgesetzt. Dieses umfasst 8 Mio. Tripel. Das LOD-Angebot des Kantons Zürich hat die möglichen 8 Mio. Tripel nicht ausgeschöpft.

Weitere Möglichkeiten zur Publikation

Kontextualisierte Daten lassen sich auch in anderer Form publizieren, in einfachster Form direkt als RDF-Dateien.

Das deutsche Konsortium [NFDI4Culture](#) hat Linked Open Data Management Services miteinander verglichen. Diese [Zusammenstellung](#) ist eine gute Basis für das Testen von Alternativen zu LINDAS.

Aufbereitungsprozess, Konvertierungsskripte, Daten-Tripel und Ontologien werden idealerweise zusätzlich auf einer geeigneten Plattform dokumentiert und der Öffentlichkeit zur Verfügung gestellt. So stehen beispielsweise die Ehedaten-Tripel der Öffentlichkeit auf [GitHub](#) weiterhin zur Verfügung, auch wenn der LINDAS-Service eingestellt wird. Sie müssen jedoch zur Abfrage in eine lokale Umgebung geladen werden, solange es keine alternative Publikationslösung gibt.

Sie verlieren damit die Vernetzung zu anderen Datensätzen und sind nicht mehr als 5-Star-Data verfügbar, sondern als 4-Star-Data. Weiter festzuhalten ist, dass bei jeder alternativen Lösung zu LINDAS der gemeinsame Endpoint der Schweizer Verwaltungen verloren geht, bzw. die LOD des Kantons Zürich nicht mehr darüber bezogen werden können.

Datenaufbereitung

Die Aufbereitung von bestehenden Daten als LOD bedeutet die Anreicherung der Daten mit Kontext. LINDAS bietet dafür mit dem Cube Creator eine einfache Möglichkeit, welche wenig Vorwissen verlangt. Allerdings ist der Cube Creator nur für die Aufbereitung in [Data-Cubes](#) geeignet. Die AG LOD hat sich mit der Verwendung des Cube Creators auseinandergesetzt und erhielt eine Einführung (vgl. [das Protokoll der AG-Sitzung vom 8. September 2022](#)).

Der Datensatz des Staatsarchivs («Ehedaten») war für die Aufbereitung mit dem Cube Creator ungeeignet und erfolgte direkt mittels Python-Script. Beide Möglichkeiten verlangen eine individuelle Bearbeitung jedes Datensatzes und lassen sich nur sehr bedingt automatisieren. Die Datenaufbereitung bleibt damit zeitaufwändig.

Das Vorgehen hat die AG LOD in Guidelines für die Publikation von LOD im Kanton Zürich beschrieben und vereinheitlicht. Die Guidelines liegen momentan ebenfalls auf dem privaten Repository der AG auf GitHub.

Geöffnete Datensätze

Im Erprobungszeitraum hat die AG einen Datensatz auf LINDAS publiziert. Es handelt sich um den Datensatz «Wissensnetz der Zürcher Ehedaten 16.-18. Jh.» des Staatsarchivs Zürich. Die Zürcher Ehedaten der frühen Neuzeit beinhalten sämtliche Eheeinträge aus den Pfarrbüchern des Zürcher Herrschaftsgebiets. Sie bieten sich als Datengrundlage für verschiedene Forschungsperspektiven an.

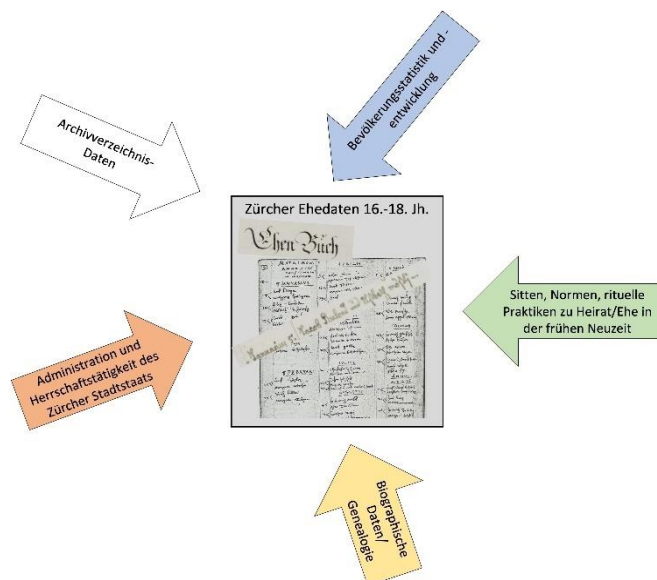


Abbildung 2: Darstellung der "Zürcher Ehedaten" des Staatsarchivs des Kantons Zürich

Die weitere Beschreibung und die RDF-Dateien mit den Daten selbst finden sich auf [GitHub](#).

Nutzungszahlen/Erfahrungen mit der Nutzung des Datensatzes

Die Linked Open Data wurden in den OGD-Katalogen verzeichnet. Die Nutzungsstatistik von opendata.swiss verzeichnete die Zugriffszahlen auf die Zürcher Ehedaten konstant in den Top 10 der Zugriffen auf OGD des Kantons Zürich. Es ist wahrscheinlich, dass die hohen Zugriffe auch auf den Bedarf an Übungsdatensätzen für Linked Data zurückzuführen ist.



Kosten

Für LINDAS, Kategorie S, während 1.5 Jahren total CHF 40'500.-.

Die Kosten für das erste Jahr wurden unter den AG-Mitgliedern aufgeteilt. Für das letzte Halbjahr hat das Statistische Amt die Kosten übernommen, abzüglich eines Anerkennungsbeitrages von 5000.- zur Projektfinanzierung des direktionsinternen Innovationsgefässes PPM JI.

Erkenntnisse und weiteres Vorgehen

Die AG LOD hat im Oktober 2023 den Versuch Linked Open Data über LINDAS zu publizieren vorläufig beendet, die Fach- und Koordinationsstelle Open Government Data hat den Vertrag mit LINDAS entsprechend auf Ende 2023 aufgelöst.

Der Entscheid im Oktober 2023 ist auf zwei Punkte zurückzuführen: die Erfahrungen mit der Publikation von Linked Data generell und die Erfahrungen mit LINDAS.

Erkenntnisse in der Publikation von Linked Open Data

- Der Versuch mit Linked Open Data hat der AG ein Gefühl für den Aufwand und die Möglichkeiten von Linked Data gegeben und erste Ansprechpersonen für Linked Data im Kanton Zürich hervorgebracht. Der Versuch ist daher positiv zu bewerten.
- Ebenfalls als äusserst positiv zu erwähnen ist der Austausch mit Statistik Stadt Zürich. Dieser war enorm wertvoll. Die Stadt und der Kanton haben in diesem Thema von Anbeginn an die Erfahrungen geteilt und der Kanton Zürich konnte von der Erfahrung der Stadt Zürich profitieren, welche Linked Open Data konsequent ins Zentrum ihrer Datenprozesse stellt.
- Die Ehedaten haben sich als sehr geeignet für die Publikation als Linked Open Data herausgestellt und waren seit ihrer Publikation beliebte OGD. Generell zeigen sich LOD als sehr interessant für Archivdaten, da die triple in ihrer Struktur statisch bleiben und die Metadaten bereits enthalten.
- Linked Data benötigt viel Zeit für den Einstieg – sowohl für die Data Publishers wie auch für die Data Users. Die Kommunikation des Wertes über einen kleinen Kreis an Linked-Data-Interessierten ist der AG nicht gelungen. Der Kreis der AG-Mitglieder ist nicht gewachsen, es gab keine weiteren Linked Open Data.
- Linked Data ist ein Nischenthema, weswegen Daten nicht systematisch verfügbar sind. Ohne eine kritische Menge an Linked Data, lässt sich die Kraft der Technologie nicht nutzen und das Momentum geht verloren.
- Innerhalb des Kantons Zürich werden Daten in zwei Projekten ¹als Triple aufbereitet und genutzt. Darüber hinaus findet kein Wissens- und Praxisaufbau statt und die

¹ Neben der im Bericht erwähnten Ehedaten, wird auch das in Umsetzung befindliche Projekt «reLex» Linked Data nutzen.



Technologie bleibt unbekannt. Daten werden tabellarisch angelegt und publiziert, was ebenfalls eine Verknüpfung zulässt. Es besteht momentan kein breiter Bedarf für eine Alternative zu LINDAS zur Publikation von LOD.

Erkenntnisse in der Nutzung von LINDAS

- Das Kostenmodell von LINDAS beruht auf der Menge Triple, die eine Organisation publiziert. Die Kosten steigen damit schnell an und verhindern eine systematische Publikation von Linked Open Data. Das Momentum für Linked Data wird dadurch effektiv durch das Preismodell von LINDAS zerstört.
- Linked Data könnte durch Reasoning angereichert werden. Reasoning erlaubte, die Daten automatisiert zu verknüpfen und damit neue Metadaten (und neues Wissen) abzuleiten. Diese Möglichkeit würde den Wert von einzelnen Linked Data und der gesamten Technologie massiv steigern. LINDAS nutzt kein Reasoning auf seinen Daten und verpasst es damit, Mehrwert zu schaffen.
- Für die Publikation von Daten gibt es aktuell noch etwas wenig Anknüpfungspunkte in Form von fixfertigen Ontologien. Auch ist die Frage, welche Ontologien sich für welche Daten eignen und welche Ontologien auf LINDAS bereits in Verwendung sind, für Nutzende von LINDAS ohne entsprechendes Expertenwissen schwierig zu beantworten. Dies mindert die Möglichkeiten der Verknüpfung der Daten.
- Der Cube Creator in Verbindung mit Visualisierungen bietet einen hohen Mehrwert und erleichtert die Publikation und Kommunikation von LOD. Er verlangt aber genau eine Ontologie, was seine Nutzung stark einschränkt. Hier wünscht sich die AG LOD einen Ausbau für weitere Ontologien (auch der gesamten Plattform LINDAS).
- Die Beratung und Begleitung der Nutzenden von LINDAS steht nicht im Fokus des Bundesarchives. Die AG LOD empfiehlt für die stärkere Verwendung von LINDAS die Ressourcen hierfür auszubauen.

Weiteres Vorgehen

- AG LOD bleibt als loses Netzwerk von Ansprechpersonen und Expert*innen bestehen.
- Die Zürcher Ehedaten sind weiterhin als RDF-Dateien auf GitHub verfügbar. Weitere Daten können analog publiziert werden. Solange der Bedarf für eine eigene Infrastruktur zur Publikation von LOD nicht gegeben ist, verzichtet die Fach- und Koordinationsstelle darauf.
- Bei erneutem Bedarf, Linked Open Data publizieren zu können, adaptiert die Fach- und Koordinationsstelle die erarbeiteten Guidelines der AG LOD auf die gegebene Situation und bietet Hand für eine neue Infrastruktur zur Publikation.
- Für das Projekt «reLex» wird zur Zeit ein Triplestore aufgesetzt. Ein SPARQL-Endpoint wird hierfür ebenfalls angeboten werden. Das Projekt hostet die Daten selbst und nutzt dazu nicht LINDAS.



Der Kanton Zürich publiziert seine Open Data dezentral. Die datenhaltenden Stellen beschreiben, publizieren und aktualisieren ihre Daten nahe am eigentlichen Verwendungszweck. Bereits in der Vergangenheit hat sich gezeigt, dass diese Organisation OGD hervorbringt, welche die Kantonsverwaltung auch selbst wiederverwenden kann. Der Bedarf für die Publikation von Linked Open Data in der Kantonsverwaltung Zürich wird erst steigen, wenn Linked Data-Technologien bereits innerhalb der Kantonalen Verwaltung für die Erfüllung der kantonalen Aufgaben genutzt werden bzw. anderweitig publizierte LOD durch die kantonale Verwaltung wiederverwendet werden können.

Zukunftsweisend könnte deswegen der Prototyp der Bundeskanzlei sein, der die Haltung von Stammdaten als Linked Data erproben möchte. Dazu gehört der Aufbau von Wissen um Linked Data und einer Infrastruktur zur Speicherung der anfallenden Daten.



Steckbrief Prototyp Linked Data

Fragestellung	Kann ein Prototyp gebaut werden, mit welchem verteilt geführte Attribute zur natürlichen Person dieser zugewiesen, dezentral verwaltet und gemeinsam genutzt werden können?
Organisation	<ul style="list-style-type: none">• Projekt Datenökosystem (DVS)• Ausführend BK-DTI
Zeitraum	Januar 2024 bis September 2024
Vorgehen	Iterativ – wir kommen so weit wie wir kommen mit dem gegebenen Budget
Zusammenarbeit	Für alle Verwaltungen der Schweiz offen Verband Schweizerischer Einwohnerdienste
Abgrenzung	Klärung der Rechtsfrage, fehlende Rechtsgrundlagen für operative Umsetzung

Abbildung 3: Prototyp Linked Data für Stammdatenmanagement der Bundeskanzlei (Quelle: Präsentation der BK, 15.02.2024)

Ebenfalls zukunftsweisend kann die Linked-Data-Nutzung bzw. Veröffentlichung durch das BFS und die Stadt Zürich sein.

Stark verteilt genutzte Daten, wie die Stammdaten, dürften mit Linked Data ideal gemanaged werden können. Die verbreitete Nutzung von Linked Data in anderen Verwaltungen und für verwaltungsnahe Datensammlungen, wird langfristig auch für die Kantonsverwaltung Zürich den Impuls geben, sich verstärkt mit der Linked-Data-Technologie zu befassen.