

Datenstrategie

Stadt Luzern

Datum: 09.02.2024
V 1.0

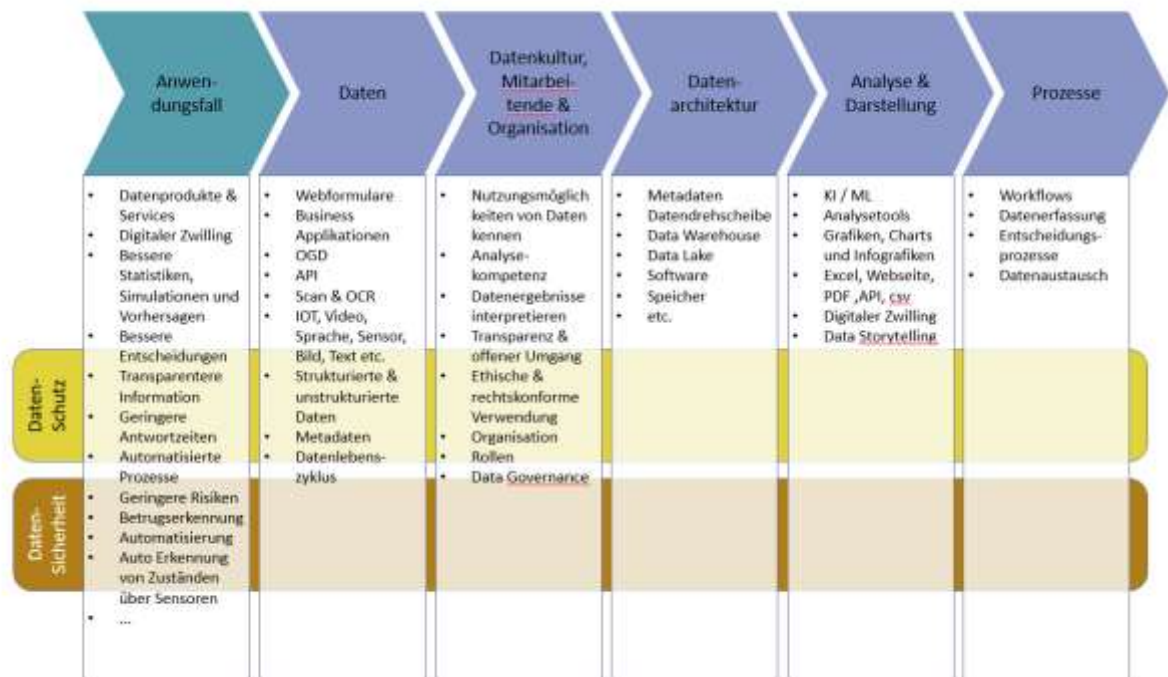
Inhaltsverzeichnis

1	Vision der Datennutzung in der Stadt Luzern	4
2	Anwendungsfälle	4
2.1	Statistische Auswertungen	4
2.2	Datennutzung zur Prozessoptimierung	4
2.3	Simulation	5
2.4	Digitaler Zwilling	5
2.5	Entwicklung von Anwendungsfällen	5
3	Daten	6
3.1	Open Government Data	6
4	Datenkultur, Mitarbeitende & Organisation	8
4.1	Datenkultur	8
4.2	Mitarbeitende	8
4.3	Organisation:	9
5	Analyse und Darstellung	11
6	Datenarchitektur	12
7	Prozesse	13
8	Datenschutz und Datensicherheit	14
9	Anhang	15
9.1	Glossar	15
9.2	Exkurs: Exemplarische Datennutzung entlang von Prozessen am Beispiel	20
9.3	Weitere Informationen	21

Präambel

Ohne Daten funktioniert heute kein Unternehmen und auch keine Verwaltung mehr. Daher kommt der Gewinnung, Verarbeitung, Speicherung und Nutzung der Daten eine überragende Bedeutung zu. Dem Verständnis zur Datennutzung bei den Mitarbeitenden kommt aus eben diesem Grund ebenfalls eine sehr hohe Bedeutung zu. Genauso müssen Organisationen heute eine Infrastruktur haben, auf der diese Datenprozesse End-to-End (von Anfang bis Ende) und medienbruchfrei abzuwickeln sind. Damit die Stadt Luzern sich auf diese Datennutzungen verstärkt ausrichten kann, gibt es diese Datenstrategie.

Die Stadt Luzern orientiert sich an der Datenwertschöpfungskette. Diese geht davon aus, dass die Daten immer einem Anwendungsfall dienen. In jedem der Bereiche der Wertschöpfungskette werden Ziele, Handlungsfelder und Massnahmen dargestellt.



Die Datenstrategie zeigt, was die Stadt in den nächsten 3-5 Jahren vorhat und wie die allgemeine Ansicht (und damit die Grundlage für Entscheidungen) zu den einzelnen Bereichen der Wertschöpfungskette ist. Nach Mintzberg ist eine Strategie ein Muster in einer Serie von Entscheidungen. Die allgemeine Kenntnis der Ansichten ist eine wichtige Grundlage zur Umsetzung der Strategie.

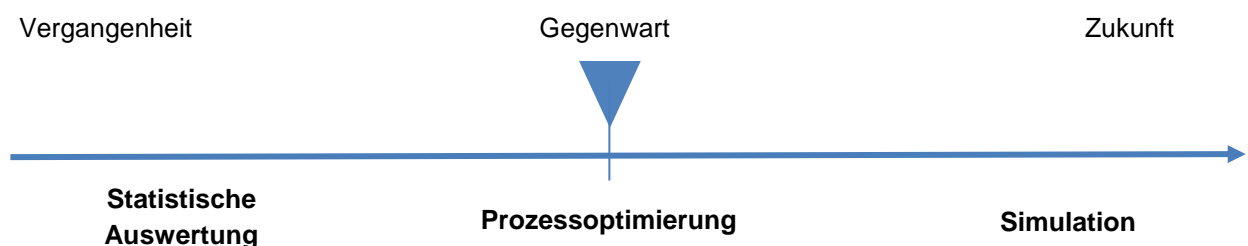
Der Text der Strategie hält sich mit Absicht an möglichst einfache Formulierungen, um zu zeigen, dass Daten ein einfaches Thema sind und keiner von den Erkenntnissen ausgeschlossen wird.

1 Vision der Datennutzung in der Stadt Luzern

Die Mitarbeitenden der Stadt Luzern verstehen, nutzen und teilen die wertvollen Daten. Damit gestalten sie die beste Stadt Luzern mit beeindruckenden Kundenerlebnissen mit.

2 Anwendungsfälle

Die Stadt Luzern unterscheidet folgende primäre Datenanwendungsfälle:



2.1 Statistische Auswertungen

Im 19. Jahrhundert hat der Schotte John Sinclair das Wort Statistik als das allgemeine Sammeln und Auswerten von Daten bezeichnet. Man beginnt die Daten in geeigneter (Technologie) und verständlicher (Metadaten) Weise abzulegen, also zu sammeln (vgl. Datenarchitektur). Von da aus werden die Daten mit verschiedensten Methoden analysiert und Erklärungsmuster für Daten gesucht. Methoden der Künstlichen Intelligenz basieren oft auf Weiterentwicklungen der Statistik.

Ziel: Mit den statistischen Auswertungen möchte die Stadt Luzern den Status der Stadt darstellen (z. B. Darstellung der Bevölkerungsstatistik über die Zeit) und Erklärungen suchen, warum die Entwicklung so war. Diese Auswertungen sollen für alle einfach zugänglich sein.

2.2 Datennutzung zur Prozessoptimierung

Die Nutzung von Daten in einem Prozess kann sehr einfach sein. In der Stadt Luzern kennen wir beispielsweise die Prüfung, ob eine antragstellende Person wirklich in Luzern wohnt. Komplexere Datenverarbeitung in Echtzeit mit KI können Sprache zu Text Anwendungen, Scannen zu Text, oder andere Mustererkennungen in Daten sein. Insgesamt werden Daten in vielfältiger Weise in Prozessen eingesetzt. Diese Datennutzung ist heute oft noch mit Medienbrüchen versehen.

Ziel: Die Stadt Luzern möchte zur Beschleunigung von Prozessen die Datennutzung so gestalten, dass die Daten jedem, der sie zur Erfüllung der Aufgabe braucht (und das kann auch eine Maschine sein), möglichst einfach zu Verfügung stehen. Zur Modernisierung von Prozessen soll die Datennutzung automatisiert werden.

2.3 Simulation

Mit den gefundenen mathematischen Zusammenhängen versuchen wir die Zustandsveränderungen durch Variation von Eingangsvariablen vorauszusagen. Das ist Simulation. Dabei können Chancen und Risiken ebenso simuliert werden, wie Finanzsituationen, Verkehrsströme, Abwasserflüsse und viele andere Zustände.

Ziel: Die Stadt Luzern möchte zur besseren Entscheidungsfindung und für mehr Transparenz Simulation in den verschiedenen Bereichen - über die heutige Nutzung hinaus - einsetzen.

2.4 Digitaler Zwilling

Alle drei Anwendungsfälle zusammen werden gerne als «Digitaler Zwilling» bezeichnet. Man denkt dann sofort an ein Mini Stadtmodell, in dem das Gleiche passiert wie in der echten Stadt und wo man frühere und zukünftige Zustände und Prozessabläufe sehen kann. In einzelnen Anwendungsfällen kann das der Fall sein und es können entsprechende Datenprodukte entstehen. Generell ist ein digitaler Zwilling eine Sammlung von Daten, die das Objekt (hier die Stadt) beschreiben. Im Bereich der Smart City ist der Digitale Zwilling ein oft auffindbares Ziel.

Ziel: Die Stadt Luzern möchte einen digitalen Zwilling – und damit ein möglichst vollständiges Datenabbild der Stadt – erarbeiten.

2.5 Entwicklung von Anwendungsfällen

Das Auffinden und die Umsetzung von Anwendungsfällen in den einzelnen Dienstabteilungen der Stadt Luzern ist in der Verantwortung der Dienstabteilungen selber. Bei Bedarf werden sie durch die Dienstabteilung Digital unterstützt.

Laufende Aufgabe: Die Dienstabteilungen nutzen Daten zur kundenorientierten Erfüllung der Aufgaben und zur Optimierung der Effizienz

3 Daten

Die Daten der Stadt Luzern werden heute primär in den Fachapplikationen der Dienstabteilungen gehalten, wo sie oftmals händisch erfasst werden. Die Hauptanwendungsfälle (oder auch Services) der Datennutzung der Dienstabteilungen sind in den einzelnen Fachapplikationen umgesetzt. Es bestehen heute viele direkte Schnittstellen zwischen den Applikationen. Sofern man diese direkten Schnittstellen zwischen den Fachapplikationen über die neue Datenarchitektur umsetzen würde und damit den digitalen Zwilling erarbeitet, wären die Daten auch für andere Anwendungsfälle einfach zu erreichen.

Ziel: Digitaler Zwilling – vollständiges Datenabbild

Handlungsfeld: Datenverzeichnis

Massnahme: Ein für alle durchsuchbares Datenverzeichnis erstellen

Wer: DIG

Bis wann: Ende 2024

Laufende Aufgabe:

- Immer wieder neue Datensätze in die Datenarchitektur der Stadt einbringen
- Abbau von festen Schnittstellen und umstellen auf die neue Datenarchitektur
- Die in die Datenarchitektur der Stadt aufgenommenen Daten entsprechend beschreiben (Metadatenkatalog).

3.1 Open Government Data

Die Stadt Luzern besitzt bereits eine OGD Strategie¹ (Open Government Data Strategie). Ziel ist es, eine Vielzahl an Daten (z. B. Baumkataster) in maschinenlesbarer Form öffentlich auf <https://opendata.swiss> zu Verfügung zu stellen. OGD Strategien verfolgen das Ziel, dass Dritte sich dieser Rohdaten bedienen können und Angebote entstehen, die die Verwaltung selber nicht im Fokus hat. Die OGD Daten der Stadt sind im Metadatenkatalog² beschrieben

Heute werden in der Stadt Luzern primär Daten mit Ortsbezug durch das Geoinformationszentrum zu Verfügung gestellt. Für andere Daten ist der Umgang und der Prozess zur Veröffentlichung noch ungeklärt. Ebenso ist es schwer etwas zu veröffentlichen, von dem man nicht weiss, dass man es hat.

Ziel: Die Stadt stellt eine umfangreiche Sammlung an Open Government Daten zu Verfügung. Sobald eine «Open by Default» Rechtsgrundlage besteht, wird diese angewendet.

Handlungsfeld: Organisation

Massnahme: Datensätze ohne Geo-Bezug: Klärung der AKV und Organisation

Wer: DIG

Bis wann: Mitte 2025

Massnahme: Umsetzung der OGD Strategie in Bezug auf Nicht-GEO-Daten

Wer: DIG

Ab wann: Mitte 2025, wenn die Organisation und AKV bestimmt sind und der Datenkatalog besteht

¹ <https://prezi.com/view/LPW90J5Vcl046dAWC6UU/>

² metacat.stadtluern.ch

Laufende Aufgabe:

- Immer wieder neue Geo-Datensätze als OGD Daten zu Verfügung stellen. Dazu sollte das GIS ein entsprechendes Budget erhalten ohne das eine Projektorganisation benötigt wird.
- Die Dienstabteilungen unterstützen die OGD Strategie aktiv und regen die Nutzung ihrer Daten aktiv an
- Die Stadt Luzern führt ein Backlog in das interessierte Stellen Wünsche für die Veröffentlichung von Daten melden können

4 Datenkultur, Mitarbeitende & Organisation

4.1 Datenkultur

Datenkultur ist immer auch Entscheidungskultur und eine Kultur zur Kundenzentriertheit. Objektive und transparente Entscheidungen brauchen entsprechende Grundlagen. Das Prinzip, dass der Kunde Daten nur einmal eingeben muss (Once-Only-Prinzip) erfordert eine entsprechende Offenheit zum Datenaustausch. Rechtsvorschriften sind wichtig bei der Beurteilung der Datennutzung, sind aber keine Ausrede, sondern können ggf. angepasst werden.

Die Qualität der Daten hat eine hohe Bedeutung, denn Analysen und Prozesse auf Basis schlechter Daten, liefern schlechte Ergebnisse).

Damit die Datennutzung auch ethischen Gesichtspunkten genügt, hat die Stadt Luzern ein Ethikboard ins Leben gerufen, welches für die Diskussion von Datenanwendungsfällen hinzugezogen werden kann.

Ziel: Die Stadt Luzern möchte Daten, moderne Datenverarbeitungsmethoden (z. B. KI) und Tools rund um Daten (z. B. Qlik Sense, ChatGPT) aktiv Nutzen und hat den Mehrwert der Datennutzung erkannt. Es gilt der Grundsatz, dass Daten innerhalb des gesetzlichen und ethischen Rahmens geteilt, erweitert, zur Verfügung gestellt und genutzt werden können und das Silodenken damit aufgebrochen wird. Eine hohe Datenqualität, die Sicherheit der Daten und eine faire und transparente Datennutzung werden gewährleistet. Mit der Datennutzung soll dem Kunden ein beeindruckendes Kundenerlebnis geboten werden. Die Mitarbeitenden fördern aktiv diese Grundsätze.

Handlungsfeld: Datenkultur

Massnahme: Erarbeitung von Werten, einer Haltung und Definition einer Vorbildfunktion zur Präzisierung der Datenkultur zusammen mit den Dienstchefs

Wer: DIG mit Dienstchefs, PA

Bis wann: Ende 2024

4.2 Mitarbeitende

Damit neue Anwendungsfälle für Daten gefunden werden können, ist die nötige Kenntnis um die Möglichkeiten der Datennutzung (Data Literacy) Voraussetzung. Mitarbeitende, die spezifische Kenntnisse brauchen, bekommen die dazu passende Ausbildung. Das ist Aufgabe der Dienstchefs zusammen mit der Dienstabteilung Personal.

Ziel: Die Mitarbeitenden der Stadt Luzern haben die Fähigkeit Möglichkeiten zur Nutzung von Daten zu erkennen, deren Nutzen abzuschätzen und setzen neue Anwendungsmöglichkeiten um. Zudem können sie Datenergebnisse, wie Visualisierungen und Analysen interpretieren

Handlungsfeld: Datenkultur

Massnahme: Durchführung von Veranstaltungen und Aktivitäten mit den Mitarbeitenden zur Vorstellung von Nutzungsmöglichkeiten und Erarbeitung von Anwendungsfällen

Wer: DIG

Bis wann: Ende 2025

Massnahme: Innerhalb des Projekts BI Strategie (siehe Analyse & Darstellung) ist der Ausbildungsbedarf für Mitarbeitende zu prüfen

Laufende Aufgabe:

- Innerhalb der Kommunikationsroadmap der Dienstabteilung Digital informiert diese regelmässig über Datenanwendungsfälle
- Die DA prüfen den spezifischen datenbezogenen Ausbildungsbedarf
- Messung der Effektivität der Massnahmen über die Anzahl von Datenanwendungsfällen
- Die Anwendungsverantwortlichen gewährleisten eine hohe Datenqualität, die Sicherheit der Daten und eine faire und transparente Datennutzung bei ihren Applikationen

4.3 Organisation:

Datennutzung, der Lebenszyklus der Daten und deren Verarbeitungsprozesse müssen organisiert werden. Aufgaben und Verantwortlichkeiten sollen klar vergeben werden und das Zusammenspiel der verschiedenen Teilnehmer aufeinander abgestimmt sein. Ebenso müssen Zugriffsberechtigungen, Sicherheitsanforderungen, Rechtsgrundlagen und die verwendete Infrastruktur geklärt sein. Das Zusammen wird Data Governance genannt. Data Governance lebt, weil sie sich mit jedem neuen Anwendungsfall verändern kann. Daher muss sie nach und nach (agil) erarbeitet und gestaltet werden.

In der Stadt Luzern gibt es heute verschiedene Stellen, die sich mit Daten auseinandersetzen

- Die Datenarchitektur und der Betrieb der Datenhaltung liegen bei ZID
- Das Geoinformationszentrum erfasst Geodaten und stellt diese auf einer selbst betriebenen Infrastruktur zu Verfügung
- Die komplexe Analyse von Daten wird durch LuStat vorgenommen
- Verschiedene Auswertungen werden durch die Finanzdirektion auf der BI Software Qlik Sense erstellt und zu Verfügung gestellt
- Die Fachstelle Datensicherheit und Privatsphäre berät und überprüft die Sicherheit der Daten in Bezug auf Richtigkeit, Verfügbarkeit und Weitergabe der Daten und berät bei Fragestellungen in Bezug auf Datenschutz.
- Die Stadtkanzlei verantwortet die öffentliche Statistik
- DIG Verantwortet die OGD Strategie
- Die einzelnen Dienstabteilungen veranlassen für verschiedenste Anwendungsfälle Auswertungen, Simulationen und Prozessoptimierungen die auch durch Dienstleister erfüllt werden können.
- In der Datenarchitektur wurden verschiedene Eigentümer von Datenentitäten und Rollen definiert.
- Die Anwendungsverantwortlichen in den DA sind für die Datensicherheit, den Datenschutz, die Qualität der Daten und für die faire und transparente Nutzung in ihren Anwendungen verantwortlich.
- AI Manager sind Brückenbauer zwischen Künstlicher Intelligenz und den Bedürfnissen der Stadtverwaltung, sowohl in Bezug auf interne Verwaltungsprozesse als auch auf die Integration der Bedürfnisse der Kundschaft. Sie fördern aktiv den zielgerichteten Einsatz von KI-Technologien.

Ziel: Die Stadt Luzern besitzt eine transparente und verständliche Data Governance. Die Aufgaben, Kompetenzen und Verantwortlichkeiten sind klar geregelt und für alle transparent.

Handlungsfeld: Organisation

Massnahme: Erarbeitung einer abgestimmten Data Governance der Stadt Luzern mit den entsprechenden Strukturen, Rollen und AKV

Wer: DIG mit Beteiligung FD, GIS, SK, ZID, FDSP

Bis wann: Ende 2025

Laufende Aufgabe: Ständige Weiterentwicklung der Data Governance nach Projektabschluss
Data Governance

Neue notwendige Ressourcen: Head of Data: 80-100%, koordiniert Data Governance, Ansprechpartner für DA und DBM in Bezug auf Daten, 50% AI Manager

5 Analyse und Darstellung

Analysekompetenzen sind heute primär bei LuStat, als ausgelagertes Kompetenzzentrum für Statistiken und beim Geoinformationszentrum vorhanden. Zudem gibt es einzelne Mitarbeitende, die mittels Qlik Sense Analysen erstellen können. Gegenüber der Bevölkerung werden Daten als OGD Daten über open-data.swiss, auf den Karten des GIS, auf der Webseite von LuStat oder auf der Webseite der Stadt Luzern dargestellt. Die zukünftige Aufgabenverteilung ist im Rahmen des Massnahmes Data Governance zu bestimmen.

LuStat darf mehr (personenbezogene) Daten miteinander verknüpfen, als die Stadt Luzern darf. Daher ist es vorteilhaft komplexere Auswertungen an LuStat zu übergeben. Zudem muss LuStat nur bezahlt werden, wenn wir einen konkreten Auftrag vergeben. Der Aufbau von Stellen in der Stadt Luzern mit datenwissenschaftlichem Hintergrund ist aus dieser Sicht nicht angedacht.

Wie interne Analysen und der Zugang auf diese, für die gesamte Stadt Luzern zu organisieren sind, ist heute noch offen. Die Verwendung von Qlik Sense für die Analysen und für die Darstellung (BI oder Business Intelligence) bietet sich an, ist aber auch eine Kostenfrage. Organisatorisch fehlt in diesem Zusammenhang noch die konkrete Anbindung der Qlik Sense Infrastruktur an die Datenarchitektur. Zudem ist das Qlik Sense Kompetenzzentrum in der Finanzdirektion verortet. Die Organisation für die Stadt Luzern und der Aufbau von Qlik Sense Kompetenzen in anderen Direktionen ist zu prüfen.

Ziel: Die Stadt Luzern besitzt ausreichende Ressourcen um die Daten zu analysieren und verständlich und attraktiv darzustellen

Handlungsfeld: Organisation

Massnahme: Erarbeitung einer BI Strategie im Rahmen der gesamten Data Governance, Nachfrage in der Verwaltung, Nutzen, Tool, Kosten, Organisation, AKV, Personalbedarf, Trainingsbedarf
Wer: FD mit Beteiligung DIG, GIS, SK, ZID, FDSP
Bis wann: Ende 2024

Neue notwendige Ressourcen: BI-Verantwortlicher 50%

6 Datenarchitektur

Die Stadt Luzern hat 2022 eine Datenarchitektur³ definiert. Diese beschreibt, wie die Daten zwischen den Endpunkten (z. B. Fachapplikationen) transferiert werden und wie die Daten als digitaler Zwilling gespeichert werden sollen. Die Datenarchitektur ist so gestaltet, dass sie für eine Vielzahl von Anwendungsfällen anwendbar ist. Separat betrachtet werden muss die Datenarchitektur des Geoinformationszentrums. Die Anschlussfähigkeit wird über vorhandene Bussysteme gewährleistet. Derzeit fehlt es an den notwendigen Ressourcen zum Aufbau und Betrieb der Datenarchitektur, für den Aufbau des digitalen Zwillings und für den Aufbau des Metadatenkatalogs (nicht Geo Daten) der Stadt Luzern.

Im Bereich der IT-Architektur ist ein IT-Architektur Gremium vorgesehen. Die Datenarchitektur ist ein Teil davon.

Der Bund schafft auf seiner Seite «einen zentralen, standardisierten Zugang im Sinne eines «Daten Hub» zu den digitalen Daten und Prozessen der gesamten Bundesverwaltung...»⁴. Dies soll über die Etablierung von «interoperablen, maschinenlesbaren und auf offenen Standards basierenden Echtzeit-Schnittstellen» erreicht und durch Governance-Strukturen begleitet werden. Die Stadt Luzern muss hier Anschlussfähig sein.

Ziel: Die Dateninfrastruktur der Stadt Luzern basiert auf einer zukunftsgerichteten, leistungsfähigen, flexibel einsetzbaren und skalierbaren Datenarchitektur

Ziel: Die Datenarchitektur der Stadt Luzern ist anschlussfähig zu Initiativen des Bundes und des Kantons Luzern

Handlungsfeld: Datenarchitektur

Massnahme: Beantragung und Einstellung der nötigen Ressourcen

Wer: ZID (oder DIG)

Bis wann: Ende 2024

Laufende Aufgabe:

- Ständige Weiterentwicklung der Datenarchitektur entlang neuer Anforderungen auf Basis von neuen Anwendungsfällen und Initiativen des Bundes und des Kantons Luzern
- Ständige Zusammenarbeit mit FDSP
- Verbindliche Terminplanung für das IT-Architekturgremium vierteljährlich
- Beobachtung der Kantonalen- und Bundes-Initiativen

Neue notwendige Ressourcen: Wir benötigen 3 FTE mit den Rollen: 1 Datenarchitekt und 2 Data Engineer. Insgesamt 300%, zzgl. Head of Data & AI Manager (Siehe oben)

³ <https://intra.stadtluzern.ch/arbeitsplatz/Seiten/ITArchitektur.aspx>

⁴ <https://www.newsd.admin.ch/newsd/message/attachments/84877.pdf>

7 Prozesse

Daten werden in Prozessen verwendet. In der Stadt Luzern werden diese Prozessabläufe oft durch Fachapplikationen unterstützt. Die Aufzeichnung von Prozessen liegt heute in der Dienstabteilung Personal. Ein Kompetenzzentrum für die Optimierung von Prozessen existiert nicht. Ebenso fehlt es an geeigneter Infrastruktur.

Automatisierte und datenbasierte Prozesse können ein Kundenerlebnis mit schnellen Antwortzeiten oder standardisierter Antworten mit hoher Qualität erzeugen.

Ziel: Die Stadt Luzern besitzt weitestgehend automatisierte und medienbruchfreie Prozesse

Handlungsfeld: Prozessoptimierung

Massnahme: Studie zum Thema Prozessoptimierung. Potenzial, Bedarf an Unterstützung, Bedarf an technischer Infrastruktur etc. -> Teilweise in Projekt Hyperautomation

Wer: DIG

Bis wann: Mitte 2025

Laufende Aufgabe:

- Die Dienstabteilungen prüfen das Prozessoptimierungspotenzial und optimieren die Prozesse

Neue notwendige Ressourcen: Business-Prozessverantwortlicher 100%, weitere Ressourcen zur Prozessoptimierung (z. B. Serviceportal) tbd

8 Datenschutz und Datensicherheit

Datenschutz und Datensicherheit sind in allen Bereichen der Datenwertschöpfungskette relevant. Die Stadt Luzern besitzt eine Weisung⁵ zum Datenschutz und zur Datensicherheit.

Ziel: Die Stadt Luzern sichert die Datennutzung gemäss der Weisung zu Datenschutz und Datensicherheit

Handlungsfeld: Datenschutz und Datensicherheit

Laufende Aufgabe:

- 1 Die Stadt Luzern hält sich an die gesetzlichen Vorgaben und orientiert sich an der Weisung zu Datenschutz- und Datensicherheit

⁵ https://intra.stadtluzern.ch/arbeitsplatz/Dokumente/Weisung_Digitale_Sicherheit_und_Privatsph%C3%A4re.pdf#search=weisung%20digitale%20sicherheit

9 Anhang

9.1 Glossar

Im Thema Daten gibt es eine Vielzahl von Begriffen. Innerhalb der Strategie wurden die für das Verständnis der Strategie nötigen Begriffe erklärt. Im Folgenden werden weitere Begriffe erklärt.

Begriff	Beschreibung
Algorithmus	Eine Reihe von wiederholbaren Schritten zur Ausführung einer bestimmten Art von Aufgabe mit Daten. Typische Algorithmen sind z.B. die Sortierung von Daten, Erkennung von Mustern innerhalb von Datenbeständen oder automatische Berechnungen von Feldern anhand von mathematischen Formeln.
Analytics	Analytik bezeichnet den Verlauf der Darstellung von Schlussfolgerungen auf Basis der Rohdaten. Sie hilft uns, aussagekräftige Daten aus der Datenmasse zu extrahieren.
Anonymisierung	Personendaten sind anonymisiert, wenn alle Angaben, welche eine Person wieder bestimmbar machen würden, unwiderruflich zerstört und der Personenbezug somit irreversibel aufgehoben wird. Dies ist dann der Fall, wenn eine Zuordnung der Daten zu einer bestimmten Person von niemandem mehr oder nur mit einem unverhältnismässigen Aufwand an Zeit, Kosten und Arbeitskraft möglich ist.
Application Programming Interface (API)	Bezeichnet eine Programmschnittstelle bei der Daten abgeholt oder eingespielt werden können.
Artificial Intelligence (AI) – Künstliche Intelligenz	Künstliche Intelligenz bezieht sich auf den Prozess der Entwicklung von intelligenten Maschinen und Software, die die Umgebung wahrnehmen und bei Bedarf die entsprechenden Massnahmen ergreifen und sogar aus diesen Massnahmen lernen können.
Big Data	Big Data ist ein Überbegriff für riesige Mengen von strukturierten oder unstrukturierten Daten. Wichtiger als die Anzahl Daten ist jedoch die Verarbeitung, denn Big Data beinhaltet auch spezielle Algorithmen um anhand von grossen Datenmengen Muster zu erkennen.
Biometrische Daten	Daten, die mit speziellen technischen Verfahren zu den physischen, physiologischen oder verhaltenstypischen Merkmalen einer natürlichen Person gewonnen werden und eine eindeutige Identifizierung dieser natürlichen Person ermöglichen oder bestätigen
Business Intelligence	Business Intelligence ist der allgemeine Begriff für die Identifikation, Extraktion und Analyse von Daten. Oftmals werden Systeme eingesetzt, in denen Self Service bei Daten betrieben werden kann\$.
Datenbank-as-a-Service	Database-as-a-Service bezieht sich auf eine Datenbank, die in der Cloud auf einer Pay-per-Use-Basis gehostet wird. Zum Beispiel bei Amazon Web Services.
Daten	Durch Aufzeichnungen, Beobachtungen, Messungen, statistische Verfahren o. Ä. gewonnene) Informationen
Data Cleansing	Data Cleansing bezieht sich auf den Prozess der Überprüfung und Überarbeitung von Daten, um Duplikate zu löschen, Fehler zu korrigieren und für Konsistenz zu sorgen.
Dark Data	Dark Data bezieht sich auf alle Daten, die von Unternehmen gesammelt und verarbeitet werden, ohne dass sie für sinnvolle Zwecke verwendet werden. Sie werden dunkel genannt, weil sie ungenutzt und unerforscht

	sind. Dazu gehören Feeds sozialer Netzwerke, Call-Center-Protokolle, Meeting-Notizen usw.
Data Governance	Die Data Governance beinhaltet die implementierten Richtlinien und Verfahren, mit denen sichergestellt wird, dass die Daten eines Unternehmens von Anfang an korrekt sind und bei der Eingabe, Speicherung, Bearbeitung, dem Zugriff und der Löschung ordnungsgemäß behandelt werden. Die Data Governance hat folgende Aufgaben: Sie stellt die Infrastruktur und Technologie bereit, richtet Prozesse und Richtlinien ein, kümmert sich um deren Pflege und benennt Personen (oder Positionen) innerhalb eines Unternehmens, die für den Umgang mit und den Schutz von bestimmten Arten von Daten sowohl berechtigt als auch verantwortlich sind.
Data Lake	Der Begriff Data Lake bezieht sich auf ein Speicher-Repository, das eine riesige Menge an Rohdaten in ihrem ursprünglichen Format aufnehmen kann. Der Data Lake eignet sich zum Speichern von unstrukturierten, sowie strukturierten Daten und unterscheidet sich damit von einem klassischen Datawarehouse, welches strukturierte Daten speichern kann. Jedem Datenelement in einem Data Lake wird ein eindeutiger Bezeichner zugewiesen und mit einer Reihe von erweiterten Metadaten-Tags versehen. Der Data Lake kann nach relevanten Daten abgefragt und ein kleinerer Satz von Daten analysiert werden, um relevante Geschäftsfrage zu beantworten.
Data Mining	Data Mining ist ein weit gefasster Begriff für die Suche nach rationalen Mustern in grossen Datensätzen. Der Zweck ist es, eine riesige Datenmenge in eine verständlichere und zusammenhängende Menge von Informationen zu verfeinern.
Data Modelling	Mit Datenmodellierung bezeichnet man in der Informatik Verfahren zur formalen Abbildung der in einem definierten Kontext relevanten Objekte mittels ihrer Attribute und Beziehungen. Hauptziel ist die eindeutige Definition und Spezifikation der in einem Informationssystem zu verwaltenden Objekte, ihrer für die Informationszwecke erforderlichen Attribute und der Zusammenhänge zwischen den Informationsobjekten, um so einen Überblick über die Datensicht des Informationssystems erhalten zu können.
Data Science	Data Science ist ein breit gefächertes und übergeordnetes Thema, das Statistik, Datenvisualisierung, Computerprogrammierung, Data Mining, maschinelles Lernen und Datenbanktechnik umfasst, um komplexe Probleme zu lösen.
Data Swamp	Ein Data Swamp entsteht für gewöhnlich aus einem Data Lake. Es handelt sich um ein Datendepot, dessen Daten aufgrund fehlender Metadaten, gebrochener Beziehungen und mangelhafter Organisation nicht mehr sinnvoll für Businessanalysen nutzbar sind.
Data Pipeline	Eine Data Pipeline ist im eigentlichen Sinne eine Software, die viele manuelle Schritte zu einem Prozess zusammenführt und den Datenfluss automatisch sicherstellt - von einer Station zur nächsten. Dabei wird zunächst definiert welche Daten, wo und wie gesammelt werden. Die Data Pipeline automatisiert die Prozessschritte wie Extraktion, Transformation, Matching, Validierung und Laden von Daten für zusätzliche Analysen und Visualisierungen.
Data Warehouse	Ein Data Warehouse (DW) ist ein digitales Speichersystem, das große Datenmengen aus vielen verschiedenen Quellen miteinander verbindet und harmonisiert. Ziel ist es, Business Intelligence (BI), Berichterstellung und Analysen zu unterstützen und gesetzliche Anforderungen zu erfüllen, damit Unternehmen ihre Daten in Erkenntnisse umwandeln und intelligente, datengestützte Entscheidungen treffen können.

Datenpipeline	Eine Datenpipeline beschreibt eine Reihe von automatisierten und wiederholbaren Prozessen zum Auffinden, Bereinigen, Umwandeln und Analysieren beliebiger Daten an ihrer Quelle. Da die Daten in der Nähe des Ortes analysiert werden, an dem sie erzeugt werden, können Fachanwender die benötigten Informationen schnell und zu geringeren Kosten für das Unternehmen analysieren und weitergeben. Datenpipelines können auch durch Technologien wie maschinelles Lernen verbessert werden, um sie schneller und effektiver zu gestalten.
Datenschutz	Datenschutz bezieht sich auf die Richtlinien und Praktiken für den Umgang mit Daten in einer Weise, die sie vor unberechtigtem Zugriff oder Offenlegung schützt. Datenschutzrichtlinien und -praktiken regeln, wie Informationen im Rahmen der Datenstrategie des Unternehmens erfasst und gespeichert werden, wie sie an Dritte weitergegeben werden dürfen und wie die gesetzlichen Bestimmungen einzuhalten sind. Datenschutz ist eine geschäftliche Notwendigkeit, mit der die Erwartungen der Kunden erfüllt und gleichzeitig die Integrität und Sicherheit der gespeicherten Informationen geschützt werden.
Datenschutzfolgenabschätzung	Überprüfung und Bewertung von Risiken und deren möglichen Folgen für die Rechte und Freiheiten der Betroffenen. Risiko = Schaden x Eintrittswahrscheinlichkeit
Datensicherheit	Bei der Datensicherheit geht es darum, Daten sicher zu machen – sicher vor unbefugten Zugriffen, Katastrophen oder Systemausfällen und gleichzeitig leicht zugänglich für autorisierte Benutzer und Anwendungen. Zu den Methoden und Tools gehören Datenverschlüsselung, Schlüsselverwaltung, Redundanz- und Sicherungsmaßnahmen sowie Zugriffskontrollen. Datensicherheit ist für Organisationen jeder Größe und Art eine Voraussetzung, um Kunden- und Unternehmensdaten vor der ständig wachsenden Bedrohung durch Datenschutzverletzungen und Datenschutzrisiken zu schützen. Redundanz und Backups sind wichtig für die Geschäftskontinuität und Datenwiederherstellung im Notfall (Disaster-Recovery).
Datensilo	Ein Datensilo ist ein umgangssprachlicher Begriff für eine Situation, in der einzelne Abteilungen oder Funktionsbereiche einer Organisation keine Daten und Informationen mit anderen Abteilungen austauschen. Diese Isolierung verhindert koordinierte Anstrengungen zur Erreichung der Ziele der Organisation und führt zu schlechter Leistung (und schlechtem Kundenservice), hohen Kosten und einer allgemeinen Unfähigkeit, auf Marktanforderungen und Veränderungen zu reagieren. Doppelte und redundante Daten lassen sich nur schwer abgleichen, was die Koordinierung von Aktivitäten und die effektive Verwaltung des Unternehmens weiter erschwert.
Datenqualität	Datenqualität ist ein unscharfer Begriff, der die Eignung und Zuverlässigkeit von Daten beschreibt. Gute, qualitativ hochwertige Daten bedeuten einfach, dass die Daten genau (wirklich repräsentativ für das, was sie beschreiben), zuverlässig (konsistent, überprüfbar, ordnungsgemäß verwaltet und geschützt) und in dem Maße vollständig sind, wie es die Benutzer und Anwendungen benötigen.
Datenvirtualisierung	Die Datenvirtualisierung bietet Organisationen eine einheitliche Sicht auf alle Daten der Organisation – über unterschiedliche Systeme und Formate hinweg – in einer virtuellen Datenschicht. Anstatt Daten zu duplizieren, belässt die Datenvirtualisierung die Daten in ihren Quellsystemen und stellt den Benutzern und Anwendungen einfach eine virtuelle Darstellung in Echtzeit zur Verfügung.

Dokumenten Intelligenz	Funktionen innerhalb der KI, die sich mit der Verwaltung, Verarbeitung und Verwendung von hohen Datenmengen in Formularen und Dokumenten befassen
Echtzeit	Der Begriff Echtzeit charakterisiert den Betrieb von Systemen, die bestimmte Ergebnisse zuverlässig innerhalb einer vorbestimmten Zeitspanne liefern können. In der Regel bemerkt die Kundschaft keine Zeitverzögerungen.
ETL	Steht für Extract, Transform und Load. Es bezeichnet ein Vorgang bei dem Daten extrahiert, umgewandelt und anschliessend an die neue Zieladresse geladen werden.
ELT	Steht für Extract, Load, Transform. Es bezeichnet ein Vorrang bei dem Daten extrahiert, dann geladen und anschliessend transformiert werden.
Gamification	Gamification ist die Anwendung von Spielelementen und digitalen Spieldesigntechniken auf Nicht-Spiel-Probleme, wie z. B. geschäftliche und soziale Herausforderungen.
Generative KI	Funktionen innerhalb von KI, die ursprüngliche Inhalte in einer Vielzahl von Formen erstellen, einschließlich natürlicher Sprache, Bild, Code und mehr.
In-Memory	In-Memory bezieht sich auf ein Datenbankmanagementsystem, das Daten im Hauptspeicher statt auf der Festplatte speichert, um eine schnellere Verarbeitung, Speicherung und das Laden der Daten zu ermöglichen.
Internet of Things (IoT)	Das Internet der Dinge (Internet of Things, IoT) ist ein Netzwerk miteinander verbundener Computergeräte, mechanischer und digitaler Maschinen, Objekte, Tiere oder Menschen, die mit eindeutigen Kennungen versehen sind und die Fähigkeit haben, Daten über ein Netzwerk zu übertragen, ohne dass eine Interaktion von Mensch zu Mensch oder Mensch zu Computer erforderlich ist.
Latenz	Latenz bezieht sich auf Verzögerungen bei der Übertragung oder Verarbeitung von Daten. Es gibt zwei Arten von Latenz, die Netzwerklatenz und die Festplattenlatenz.
Machine Learning	Maschinelles Lernen ist die Entwicklung von Systemen, die auf der Grundlage der eingespeisten Daten lernen, sich anpassen und verbessern können. Dies beinhaltet typischerweise die Implementierung von prädiktiven und statistischen Algorithmen, die kontinuierlich das "richtige" Verhalten und die richtigen Erkenntnisse finden können, wenn mehr Daten durch das System fließen.
Maschinelles Sehen	Funktionen innerhalb von KI, um die Welt visuell durch Kameras, Videos und Bilder zu interpretieren.
Meta Data	Metadaten können als die Daten definiert werden, die dazu dienen, Kontext oder zusätzliche Informationen über andere Daten bereitzustellen. Zum Beispiel stellen Informationen über den Titel, das Thema, den Autor, die Schriftart, die Erweiterungen und die Größe der Datendatei eines Dokuments Metadaten über dieses Dokument dar. Sie können auch die Bedingungen beschreiben, unter denen die in einer Datenbank gespeicherten Daten erfasst wurden, ihre Genauigkeit, das Datum, die Uhrzeit, eine Methode der Zusammenstellung und Verarbeitung usw.
Natural Languages Processing (NLP) – Linguistische Datenverarbeitung	Die Abkürzung NLP steht für Natural Language Processing und beschreibt Techniken und Methoden zur maschinellen Verarbeitung natürlicher Sprache. Ziel ist eine direkte Kommunikation zwischen Mensch und Computer auf Basis der natürlichen Sprache.
Normalisierung	Normalisierung beschreibt den Prozess der Strukturierung und Umorganisation eines relationalen Datenbankschemas. Informationen werden in

	mehrere Tabellen aufgeteilt und über Beziehungen mit dem Ziel miteinander verknüpft, Redundanzen, Inkonsistenzen und Anomalien zu vermeiden und beseitigen.
Personendaten	<p>Personendaten sind Informationen, die sich auf eine bestimmte oder bestimmbare Person beziehen. Eine Person ist bestimmt, wenn sich ihre Identität unmittelbar aus den Daten selbst ergibt (z. B. Name, Adresse). Eine Person ist bestimmbar, wenn sich ihre Identität aus dem Kontext der Daten durch Kombination mit anderen Daten ergibt. Die Bestimmbarkeit liegt nicht bei unverhältnismässigem Aufwand vor.</p> <p>Zu den besonders schützenswerten Personendaten (besondere Gefahr einer Persönlichkeitsverletzung) gehören: die religiösen, weltanschaulichen, politischen oder gewerkschaftlichen Ansichten oder Tätigkeiten, die Gesundheit, die Intimsphäre, die Rassenzugehörigkeit oder die ethnische Herkunft wie Angaben zu Krankheiten, medizinischen Befunden, Medikamenteneinnahmen; sowie persönliche Angaben, die nur wenigen anvertraut werden; Zugehörigkeit zu einer Kulturgemeinschaft (z. B. Tamilen, Kurden, Sintis); Massnahmen der sozialen Hilfe, administrative oder strafrechtliche Verfolgungen oder Sanktionen</p>
Pseudonymisierung	Eine Pseudonymisierung der Personendaten liegt vor, wenn die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können. Alle Daten, die Rückschlüsse auf die betroffene Person ermöglichen, werden durch neutrale Angaben (Pseudonym) ersetzt.
Relationales Datenmodell	Ein relationales Datenmodell beinhaltet die Verwendung von Datentabellen, die Gruppen von Elementen in Beziehungen zusammenfassen. Diese Modelle beruhen auf der Idee, dass jede Tabelleneinrichtung einen Primärschlüssel oder Bezeichner enthält. Andere Tabellen verwenden diesen Bezeichner, um "relationale" Datenverknüpfungen und Ergebnisse bereitzustellen. Datenbankadministratoren verwenden Structured Query Language (SQL), um Datenelemente aus einer relationalen Datenbank abzurufen.
SQL-Datenbank	Eine SQL-Datenbank ist eine relationale Datenbank, die Daten in Tabellen und Zeilen speichert. Datenelemente (Zeilen) werden auf der Grundlage von Gemeinsamkeiten verknüpft, um die Effizienz zu steigern, Redundanz zu vermeiden und einen einfachen und flexiblen Abruf zu ermöglichen. SQL steht für Structured Query Language (strukturierte Abfragesprache).
NoSQL-Datenbank	NoSQL-Datenbanken wurden für den Umgang mit unstrukturierten Daten entwickelt, die SQL aufgrund der fehlenden Struktur nicht unterstützen kann. NoSQL verwendet kreative Techniken, um diese Einschränkung zu überwinden, einschließlich dynamischer Schemata und verschiedener Vorverarbeitungstechniken. Die gebräuchlichsten Arten von Datenbanken für unstrukturierte Daten sind Schlüssel-Wert-, Dokument-, Spalten- und Diagramm-Datenbanken und umfassen oft Objekte wie Videos, Grafiken, Freitext und Rohdaten von Sensoren
Stream Processing	Dies ermöglicht eine Echtzeitanalyse der Daten, die in das System eingespeist werden. Es ist nützlich für zeitkritische Operationen mit hoher Geschwindigkeit.
Strukturierte Daten	Sind Daten, die maschinenlesbar sind, z.B. eine relationale Datenbank, Tabellen, CSV etc.
Unstrukturierte Daten	Sind Daten, die in einem nicht maschinenlesbaren Format vorliegen, z.B. Text.

Taxonomie	Unter Taxonomie versteht man die Klassifizierung von Daten nach einem vorgegebenen System mit einem daraus resultierenden Katalog. Sie bietet einen konzeptionellen Rahmen für einen einfachen Zugriff und Abruf.
Wissensgewinnung (Knowledge Mining)	Funktionen innerhalb von KI, um Informationen aus großen Mengen meist unstrukturierter Daten zu extrahieren, um einen durchsuchbaren Wissensspeicher zu erstellen.

9.2 Exkurs: Exemplarische Datennutzung entlang von Prozessen am Beispiel

Jemand braucht eine Wohnsitzbestätigung und ruft an:

Es müssen verschiedene Systeme miteinander in Einklang gebracht werden

Anruf		
Spracherkennung und Umwandlung in Text	Speech to Text und Text to Speech - KI	«Ich brauche eine Wohnsitzbestätigung für die SBB»
Keyworderkennung oder Inhaltserkennung	Keyword oder KI	Wohnsitzbestätigung / SBB
Dialog führen	Chatbot – geführte Dialoge (hier eher kein LLM) Angabe von Name, Adresse, etc.	«Ich heiße Martina Mustermann und wohne in der Musterstr 1 in 6000 Luzern»
Formular ausfüllen	Formularsoftware wird über eingegebene Daten und / oder über API mit Daten gefüllt	
Ggf. ID Hochladen	Bild zu Text Erkennung KI	
Automatisierte Entscheidung	Adresse existiert und andere Voraussetzungen werden erfüllt. Ggf. müssen andere Systeme Daten für die Entscheidung liefern, Hier z. B. Innosolve City	
Check Out	Zahlung und Geschäftserfassung	Bitte geben Sie die Nummer 12345 in Ihrer Twint App an
Dokument erstellen und Signieren	Signatursoftware, Dokumentenengine	«Wir haben Ihnen die Wohnsitzbestätigung in Ihr elektronisches Postfach gelegt»

Jemand, der nur Englisch spricht, möchte wissen, was in einem bestimmten B+A steht

Ein solcher Fall kann heute komplett mit einer auf Large Language Modelle basierenden Systemen abgewickelt werden

Suchanfrage auf der Webseite der Stadt Luzern		
Texteingabe in Suchfeld	Automatische Spracherkennung und Übersetzung – KI – LLM	«I need a Summary of the B&A 22»
Keyworderkennung oder Inhaltserkennung	Keyword oder KI - LLM	Summary
Dialog führen	Chatbot – LLM Modell Ausgabe des Ergebnisses. Ggf. kann dieses durch Angabe der	The Summary should be 1000 Words long and I need it in German

	Länge oder einer anderen Sprache angepasst werden	
Ergebnisabgabe		Ausgabe einer 1000 Wörter langen Zusammenfassung in Deutsch

9.3 Weitere Informationen

IT Strategie Stadt Luzern: <https://intra.stadtluzern.ch/arbeitsplatz/Seiten/IT-Strategie.aspx>

GIS Strategie Stadt Luzern: <https://intra.stadtluzern.ch/arbeitsplatz/Seiten/strategie-gis.aspx>

Bericht und Antrag 29/2021 «Digitalstrategie und Smart City Luzern»: <https://www.stadtluzern.ch/aktuelles/newslist/1352167>

Datenpolitik des Bundes: <https://www.bakom.admin.ch/bakom/de/home/digital-und-internet/strategie-digitale-schweiz/datenpolitik.html>

OGD Strategie des Bundes: <https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/ogd/strategie.html>

Bund: Ziele für den Einsatz von Datenwissenschaft: <https://www.bfs.admin.ch/asset/de/23768220>

Datenstrategie Kanton Basel-Stadt: <https://www.statistik.bs.ch/dam/jcr:7723f4d5-f547-48d2-9ca8-94ff1633d0bd/Datenstrategie-Basel-Stadt.pdf>

OGD Strategie Stadt Zürich: https://www.stadt-zuerich.ch/content/dam/stzh/portal/Deutsch/Stadtrat%20%26%20Stadtpraesident/Publikationen%20und%20Broschueren/Stadtratsbeschlusse/2021/Jul/1%20STRB%20Nr.%200743_2021%2C%20Beilage%201%2C%20OGD-Strategie.pdf

Datenstrategie der Europäischen Union: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

Datenstrategie Bundesrepublik Deutschland: https://bmdv.bund.de/SharedDocs/DE/Anlage/K/nationale-datenstrategie.pdf?__blob=publicationFile