Von den Rohdaten zum Business Case auf LINDAS

Linked Data Day 2026
9. November 2026

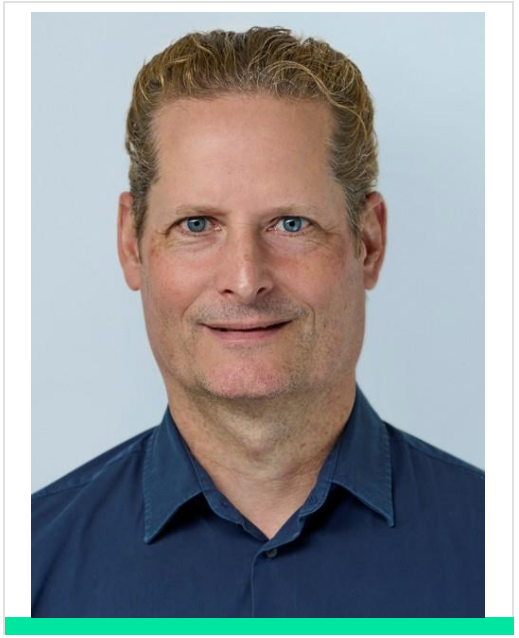
Thomas Meyhack
David Freiburghaus



adnovum

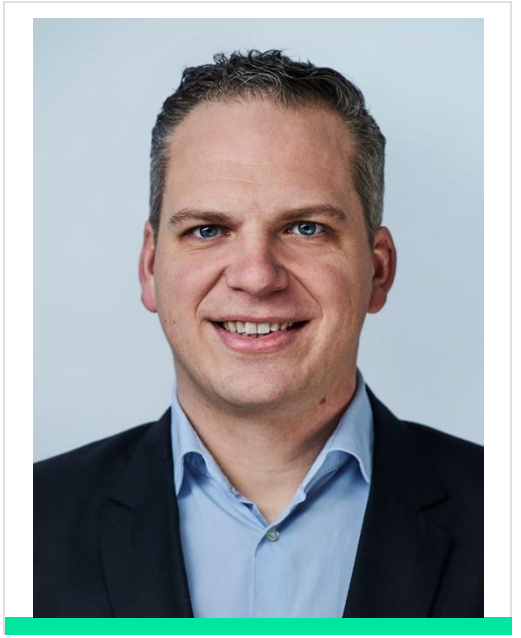
adnovum.com

Wer sind wir



Thomas Meyhack

EXPERT SOFTWARE ENGINEER



David Freiburghaus

DATA ARCHITECT

Umfrage zum Einstieg

Wer hat Erfahrung Daten von einem Triplestore zu lesen?

Wer hat Erfahrung Daten in einen Triplestore zu laden?

Wer hat praktische Erfahrung mit Cube Creator?

Wer hat praktische Erfahrung mit Visualize?



Warum sollen die Daten auf LINDAS geladen werden?

Grundlagen

- EMBAG und Open Government Data (OGD) Strategie als Vorgaben
- Once Only Prinzip: Mehrfache Nutzung derselben Daten
- Data Ownership

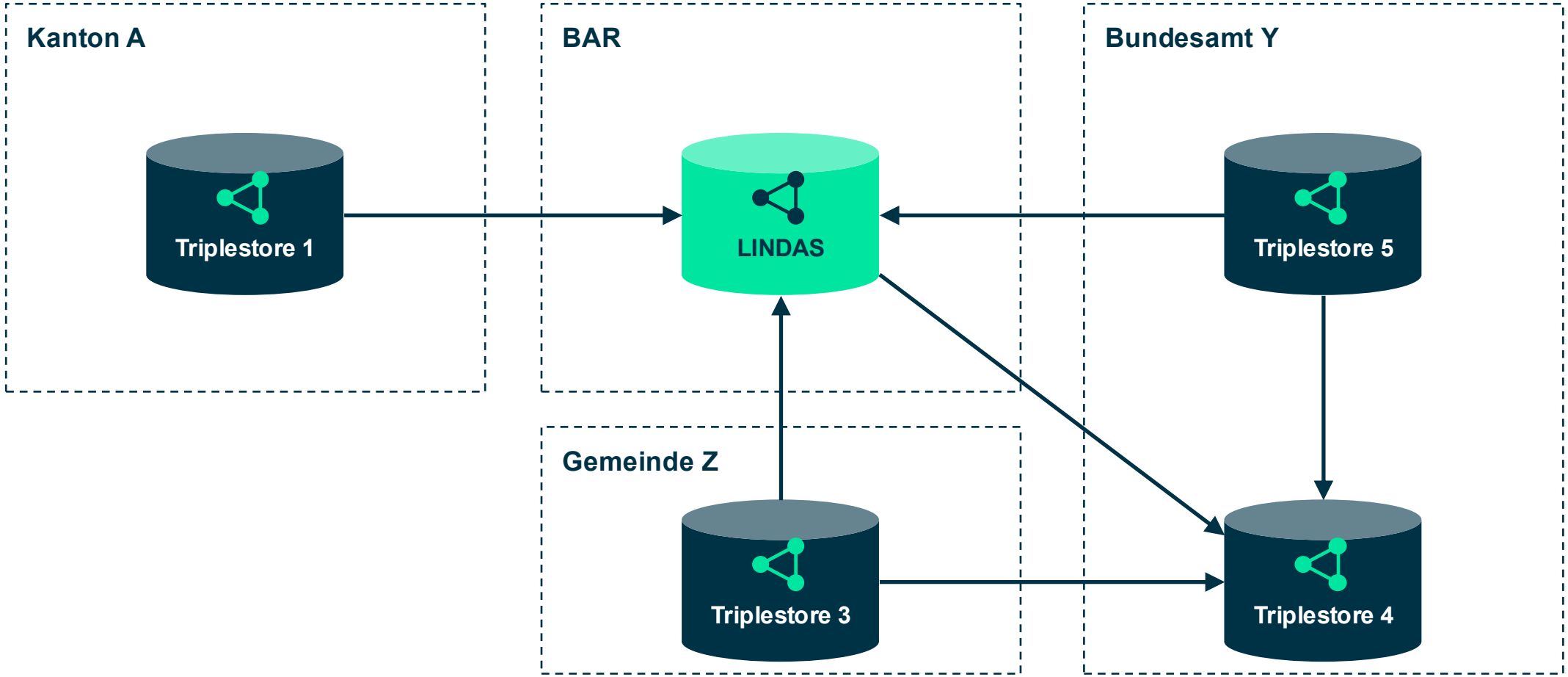
Verknüpfbare Daten

- RDF als W3C-Standard
- Stabile Identifikatoren
- Verknüpfung mit Geodaten
- Dezentrale Datenhaltung in Triplestores

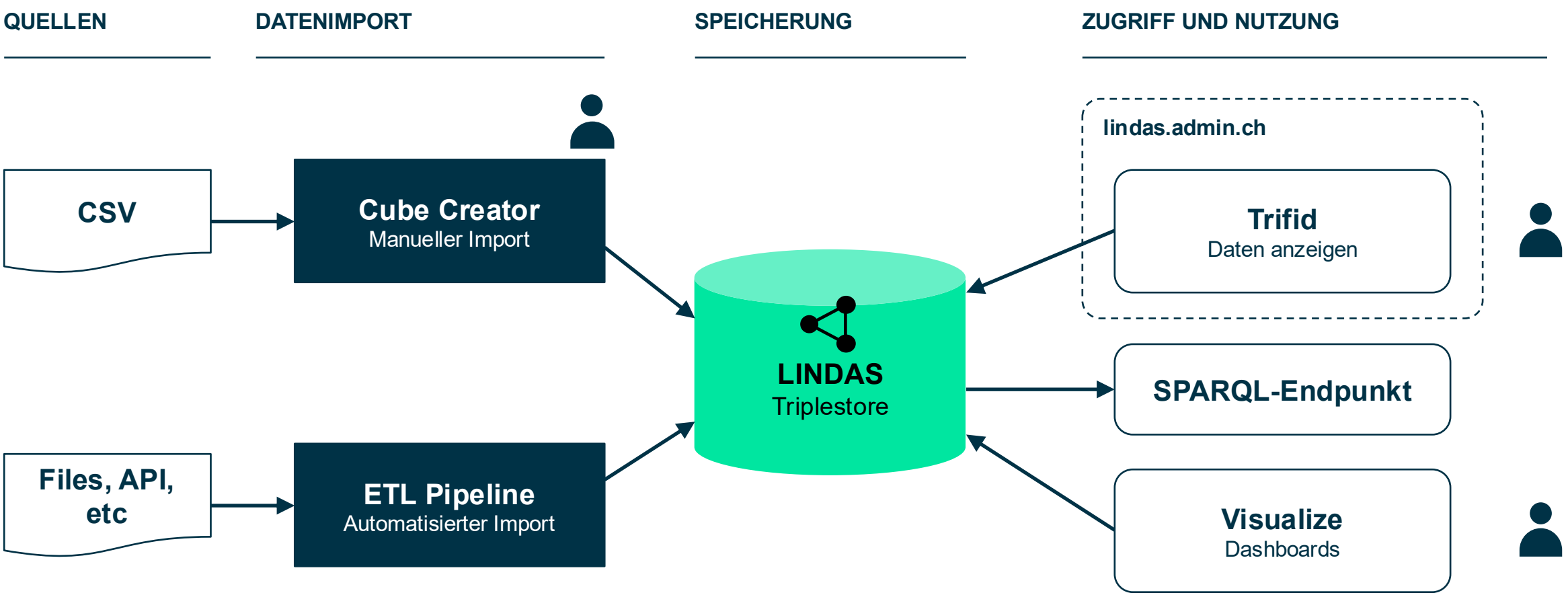
Datennutzung

- Dashboards, Integration Visualize
- Maschinenlesbare Graphen, für KI-Anwendungen
- Vereinfachte Datenarchivierung

Linked Data: OGD mit Graphdatenbanken

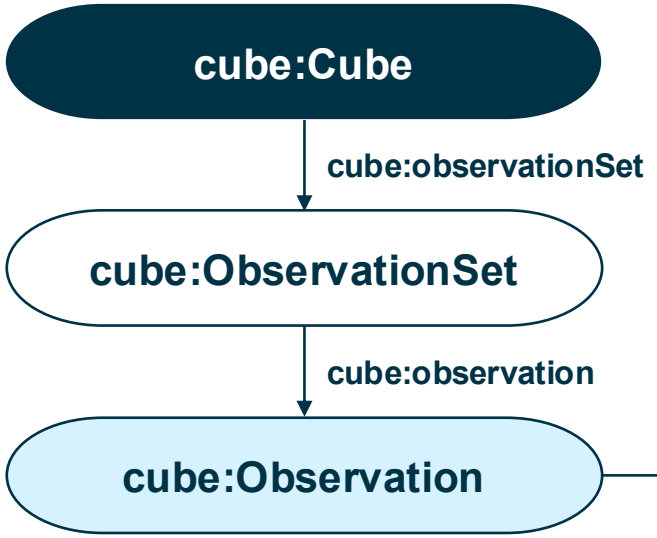


LINDAS: Systemübersicht



cube.link: Data Cubes in Triplestores

Ein minimales RDF-Vokabular für multidimensionale Datensätze (analog Tabellen)



cube:KeyDimension identifizieren gemeinsam eine Beobachtung		**cube:MeasureDimension** der beobachtete Wert
Gemeinde	**Jahr**	**Bevölkerung**
Gemeinde A	2024	12’500
Gemeinde A	2025	12’780
Gemeinde B	2025	8’340

Jede Zeile ist eine **Beobachtung**. Jede Spalte ist eine **Dimension**: eine Eigenschaft, deren Wert eine RDF-Ressource oder ein Literal ist.

Data Cubes sind optional

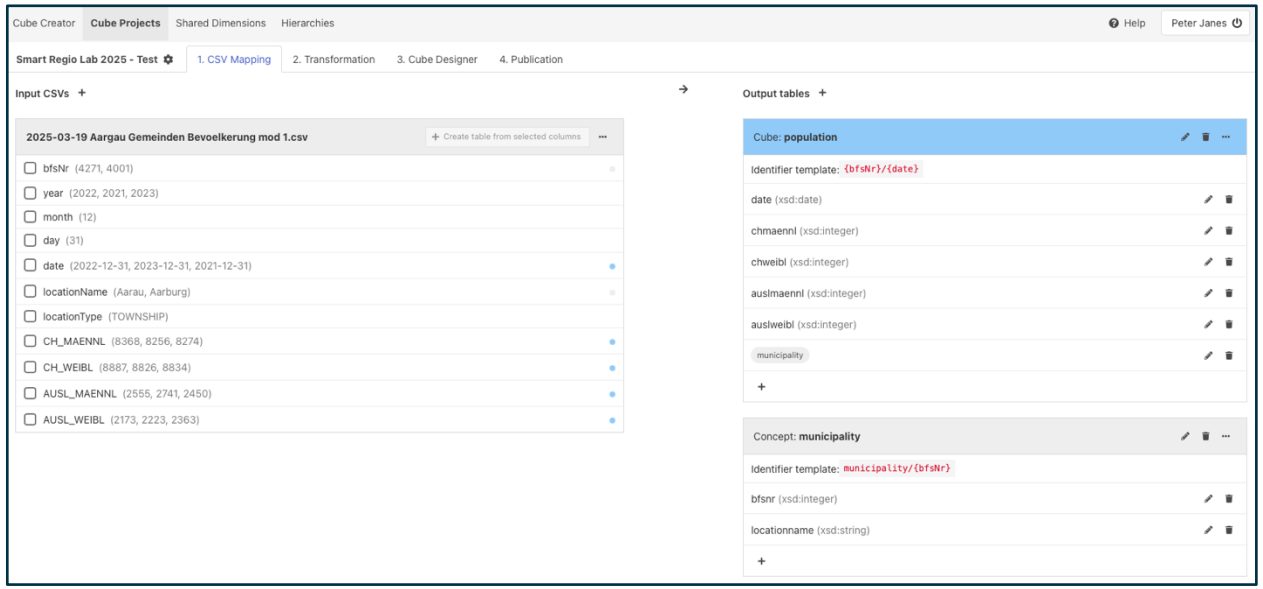
LINDAS und andere Triplestores können Cube- und Nicht-Cube-RDF-Daten gemeinsam halten, mit jeweils passenden Vokabularen.

https://cube.link/

Screenshots Cube Creator

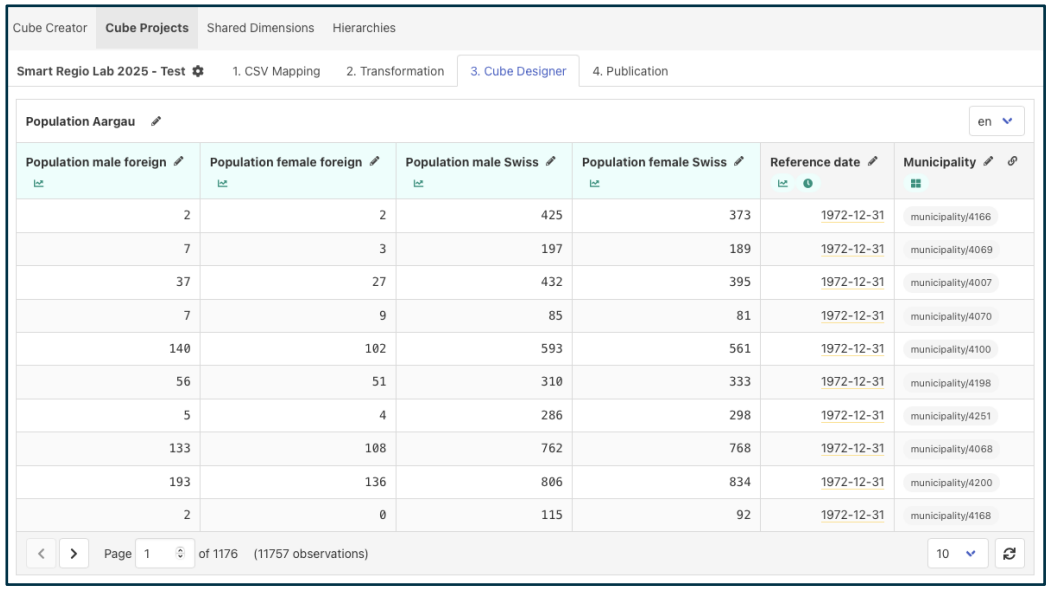
CSV Mapping

CSV-Spalten auf Cube und Konzepte abbilden



Cube Designer

Dimensionen benennen und Datentypen prüfen



Cube Creator: Features



Dient als Einstieg für tabellarische Daten

Cube Creator unterstützt die Abbildung auf cube.link

Dimensionen, Datentypen und Labels fachlich festlegen

Filterung und Formatkorrekturen

Limitierungen

Keine komplexe Bereinigung und Geschäftslogik

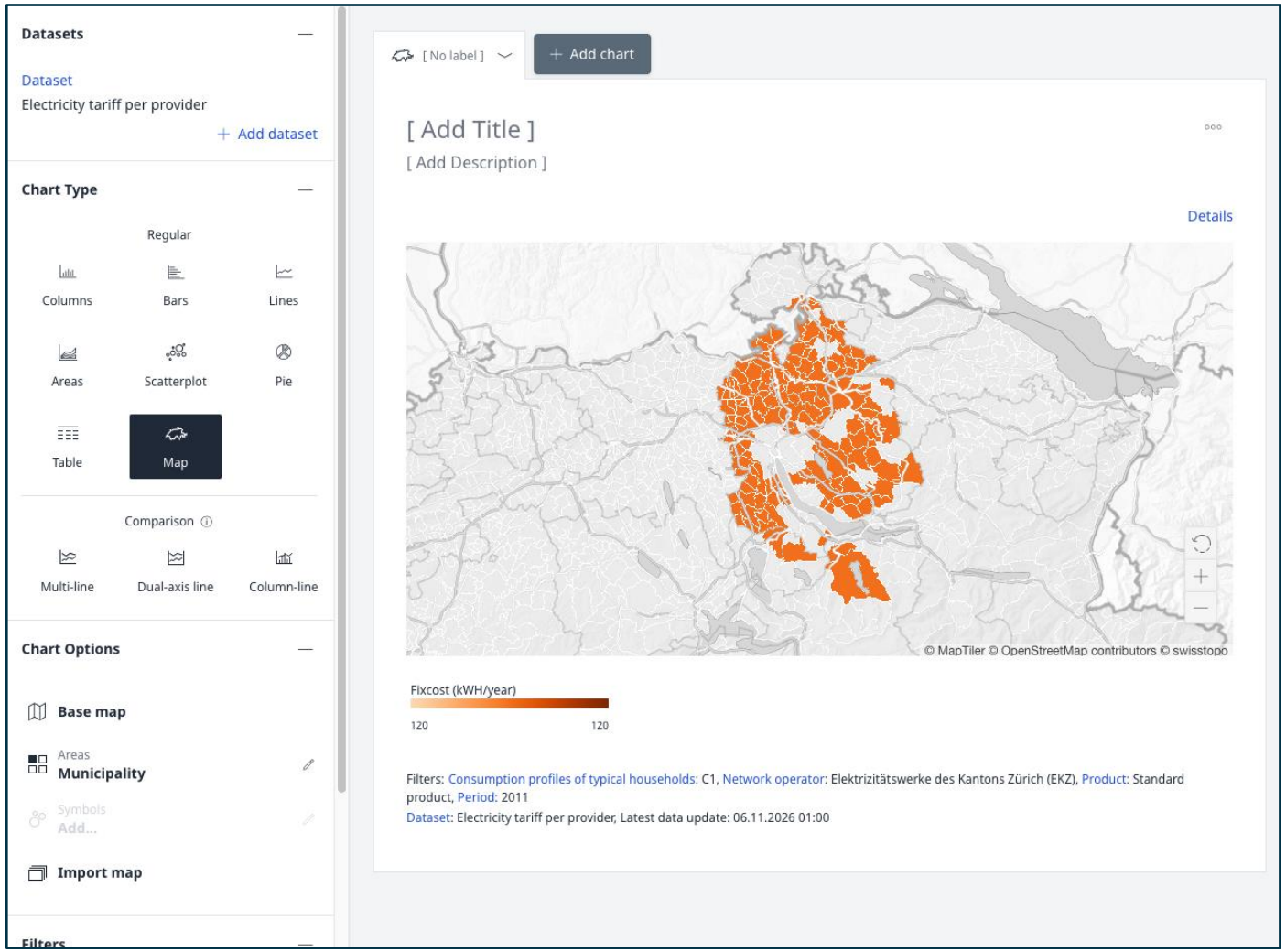
Keine Aggregationen

Beschränkt auf CSV-Input

Erstellen von RDF Cubes aus CSV Dateien via Cube Creator

https://youtu.be/r0sxnvOFqK8 (zazuko GmbH)

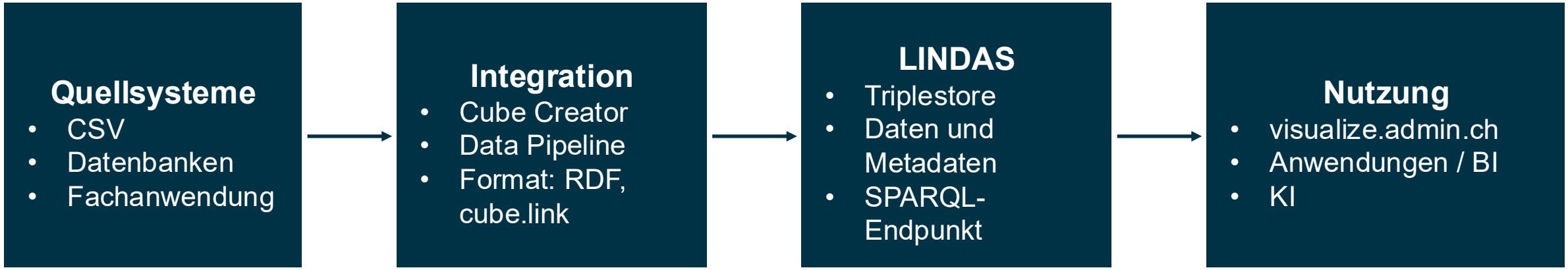
Screenshot Visualize



visualize.admin.ch
Cube aus LINDAS auswählen und als Diagramm darstellen, hier als Karte auf Gemeindeebene.

LINDAS: Daten publizieren





Was sind die Herausforderungen?



Datenverarbeitung

Daten bereinigen, modellieren und publizieren

Schützenswerte Daten (Hidden Data)

Know-how

RDF, SPARQL und LINDAS-Tooling

Aufwand und Kosten

Umsetzung des Projektes, Daten laden, Adaptieren Datenquelle

Betrieb und Wartung

Data Pipelines, Datenaktualisierung und Security

Performance

Leistung der Systeme kontinuierlich optimieren

Definition Business Case



Welche konkrete Frage soll die Daten Publikation beantworten?

Wer nutzt die publizierten Daten in welcher Form?

Welcher Nutzen zählt: einfacher Upload, Verfügbarkeit zur Auswertung, Aktualität der Daten, etc?

Datenquellen und Datenqualität



Wo liegen die Daten?

Quelle und Verantwortung
File, API, DB, etc

Format und Zugriff
Dateiformat, Schema und Zugriff klären

Historie
Verfügbarkeit historischer Daten prüfen

Volumen
Anzahl Datensätze und Grösse

Anforderungen an die Datenqualität

Integrität
Inhalt der Daten verstehen

Eindeutigkeit
Erkennen und Verarbeiten von Duplikaten

Vollständigkeit
Alle benötigten Datensätze sind verfügbar

Konsistenz
Konsistente Codes, Datentypen und Labels über Datensätze hinweg

Format
Validierung SHACL

Aktualität
Benötigtes Ladeintervall festhalten

Transformationen: Aggregationen, Joins, Filter, Business Logik

In Visualize	In einer Custom Pipeline	Im Datenmodell
• Erstellen von Dashboards • Filter und Aggregationen • Joins mit bestehenden Cubes • Performance beachten	• Komplexe Joins berechnen • Aggregation vor Publikation des Cubes • Logik versionieren und testen	• Separater Aggregat-Cube • Vorhandenen Cube erweitern mit Aggregationen, Granularität klar abgrenzen



Performance:
Ein Dashboard ist langsam. Wie gehen wir vor?



Langsame Dashboards: Diagnose und Optionen



Verarbeitungsschritte messen: SPARQL-Abfrage, Datentransfer, Darstellung

Daten vor-verarbeiten	→	**LINDAS Triplestore**	→	SPARQL-Endpunkt Datenzugriff	→	Visualize Dashboard / Client
Aggregierte Daten vorab berechnen und als Cube publizieren (vorgängig oder in Data Pipeline)		Serverseitige Verarbeitung auf Triplestore optimieren		Abfragen gezielt filtern und Datenvolumen reduzieren		Client-Cache mit klaren Aktualisierungsregeln prüfen Alternative Visualisierung prüfen (z.B. BI-Tool)

Hidden Data



Schützenswerte Daten	Bedarf
Daten mit eingeschränktem Zugriff	**Access Control**
GDPR-relevante Daten	Cube-basiert
Vertrauliche Daten	Triple-basiert
	Integration IAM System
	User und Rollenmanagement

-> LINDAS als OGD-Plattform bietet dies nicht out-of-the-box



adnovum