

Verträge für den Austausch von Daten mit GenAI erstellen und validieren

11. September 2024

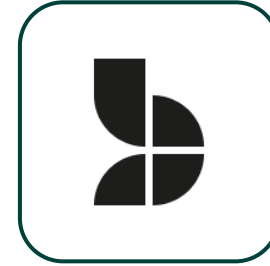
Lukas Hohl, Co-Founder & CPO

Vorstellung



Lukas Hohl

- Co-Founder & Chief Product Officer @ Beyond Civic
- Head of Blockchain @ Swisscom
- 15+ Jahre Tech Consulting in Financial Services (CH, Singapore, Boston/NYC)
- Program for Leadership Development (PLD) @ Harvard Business School
- Master of Science in Electrical Engineering & Information Technology @ ETH Zürich



Beyond Civic

- Schweizer Startup aus Luzern
- Gegründet im 2022
- Fokus auf Datenkollaborationen im Bereich von Public-Private Datenökosystemen
- Industrie: Smart Cities & Communities mit Use Cases im Bereich Nachhaltigkeit, Energie, Mobilität und Landwirtschaft
- B2B SaaS & begleitende Beratung
- Aktiv in der internationalen "Data Space Community"

Wieso sollen Daten überhaupt ausgetauscht werden?

(Wirklich) komplexe Probleme zu lösen

- Wirklich komplexe Probleme können wir nur lösen, wenn wir Daten von unterschiedlichen Organisationen zusammenbringen
- Sektor- und Länder-übergreifende Datennutzung
- Sekundärnutzung von Daten
- Verstehen von Daten durch Visualisierung

Innovative Produkte & Services zu kreieren

- Daten aus unterschiedlichen Quellen sind die Basis von neuen Produkten & Services
- Neue Business Modell ("Data as a product")
- GenAI und Digital Twins als Beispiele für die (zunehmende) Relevanz von Daten

Regulatorische Anforderungen zu erfüllen

- EU Data Governance Act und EU Data Act
- Motion "Rahmengesetz für die Sekundärnutzung von Daten" (22.3890)
- B2G Datenaustausch bei Krisensituationen
- Reporting von Scope 3 Emissionen

Risiken im Zusammenhang mit dem Austausch von Daten

Kommerzieller Aspekt

Risiko des Verlusts von Wettbewerbsvorteilen auf dem Markt.

Reputation & Vertrauen

Risiko, durch die Weitergabe oder Verwendung von Daten, das Vertrauen zu verlieren oder durch die Aufdeckung von Limitierungen in Prozessen oder Analysen einen Imageschaden zu erleiden.

Ethische Bedenken & Gesellschaft

Risiko, unethische Datenerfassung oder -verwendung zu ermöglichen oder direkte Auswirkungen auf Menschen und Gemeinschaften zu haben.

Regulatorien & Gesetze

Risiken der Verletzung von Datenschutzgesetzen, Rechten an geistigem Eigentum, anderen rechtlichen Anforderungen oder rechtlich bindenden Verträgen.

→ Gründe für das “Nicht-Teilen” von Daten sind nicht technologischer Natur

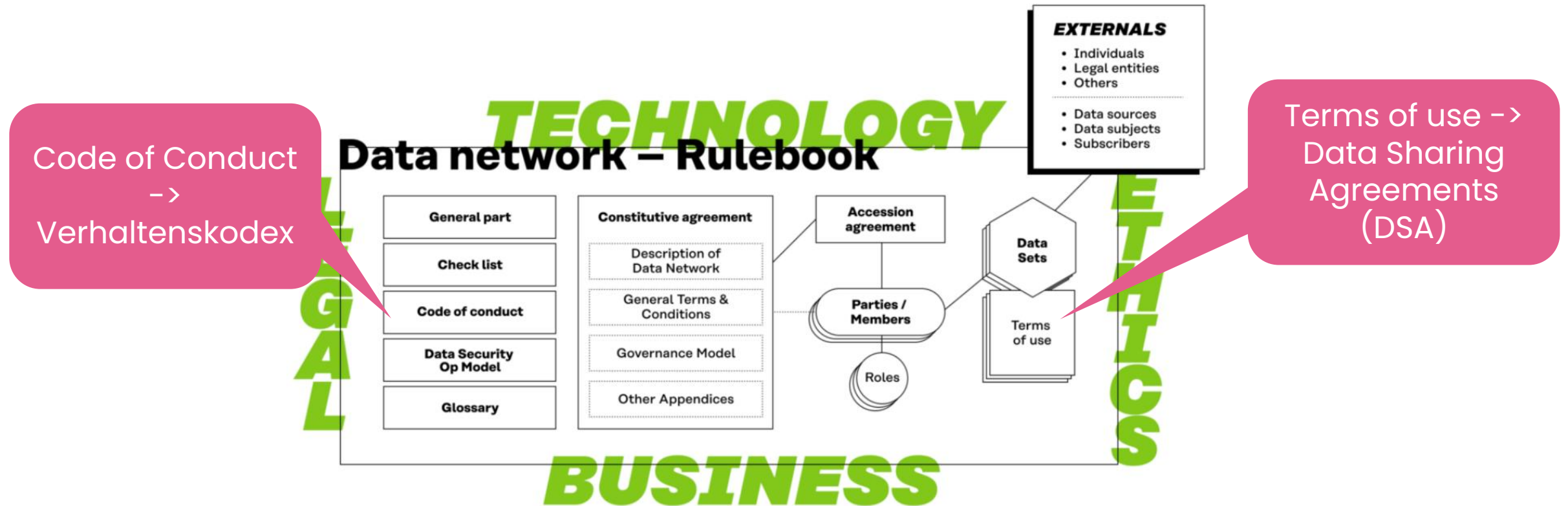
Das Versprechen von Datenräumen (Data Spaces)

Datenräume sind organisatorische und technische Strukturen, welche Datennutzende und Datenanbietende miteinander verbinden und dadurch die Mehrfachnutzung von Daten ermöglichen.

- Ein Datenraum ist eben nicht nur eine technische Infrastruktur sondern definiert auch Spielregeln (“Rulebook”), welche zusätzliches Vertrauen bringen
- Es gibt nicht DIE eine Definition was ein Datenraum ist
- Die Entwicklungen sind sehr stark getrieben durch die Europäische Datenstrategie

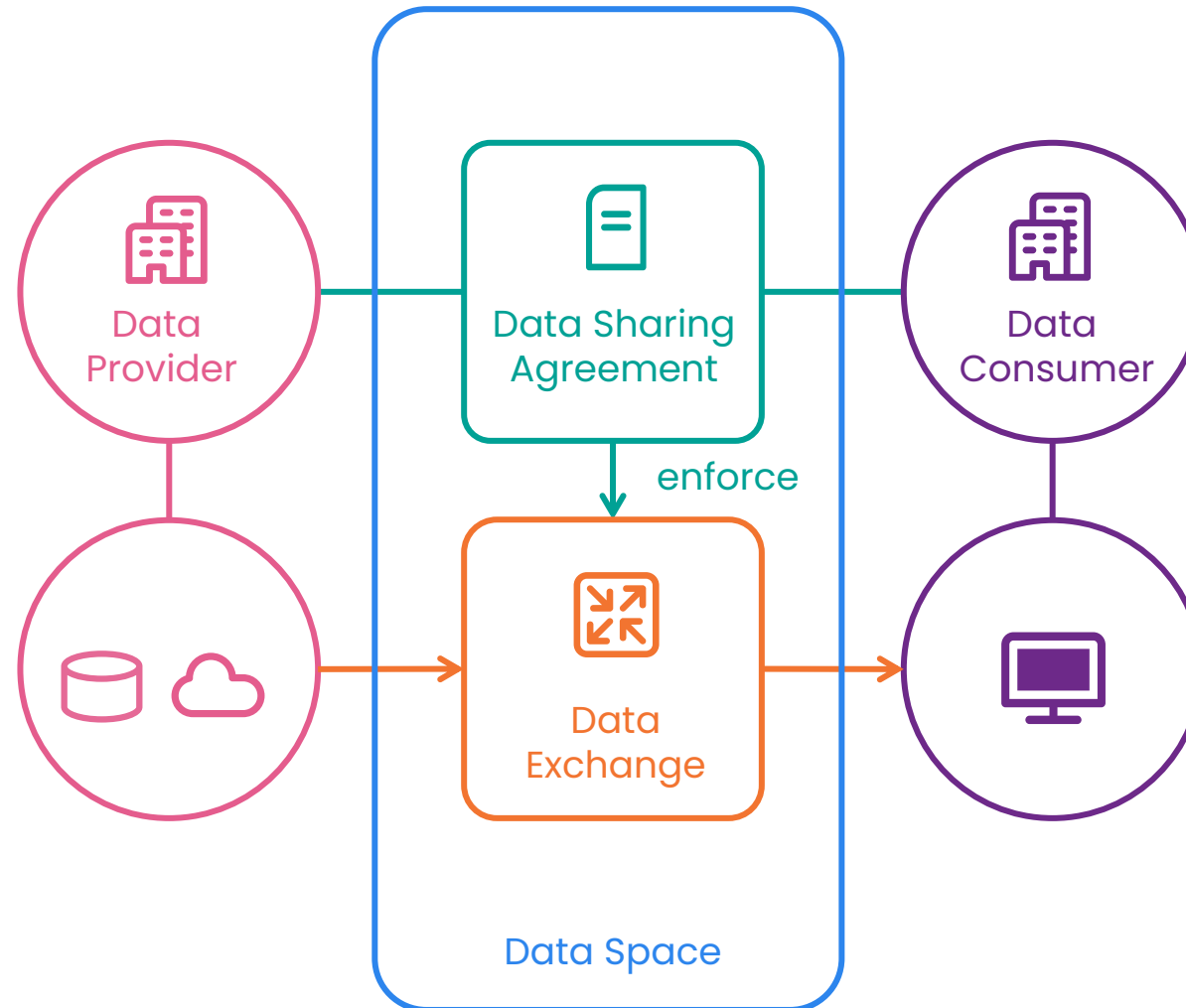
Datenraum "Rulebook"

- Sitra und International Data Space Association (IDSA) haben die Definition des Rulebooks geprägt



Quelle: [Rulebook for fair data economy \(Sitra, 2022\)](#)

Anwendung von Data Sharing Agreements im Kontext eines Datenraums



Verwendung der DSA Templates vom IGE

Create Data Sharing Agreement
Swiss IPI Template

Customize your agreement here:

Title
Sample Data Sharing Agreement

The title of the contract. Cannot be empty

When is the contract duration defined?

and access in days. Cannot be empty.

Duration until
2024-12-05

What is the purpose of the agreement?

Research

Include definitions in the agreement?

Yes

Covered Data

The data covered by the agreement, e.g. inventory records 2023. Cannot be empty.

What is the covered data format?

Completion 50%

Continue

Preview:

Sample Data Sharing Agreement

Cover sheet

Data access period	05/09/2024 - 05/12/2024
Data provider	<i>[Dynamic]</i>
Data consumer	<i>[Dynamic]</i>
Covered Data	
Purpose	Research
Notice period days	30
Data retention days	30
Requires fees	No
Subscription fee	N.A.
Usage restrictions	
Availability Service Level Agreement	N.A.
Electronic signature	None

Definitions

Commercial Use	Any use of the Covered Data and/or of the Results by the Data User or a Downward Recipient that is primarily intended towards a monetary compensation or commercial advantage.
Confidential information	Any proprietary and/or non-public information provided by either party, in particular but not limited to trade secrets or know-how or other related proprietary business information and data, whether such information is provided in tangible or intangible form, written, oral, graphic, pictorial or recorded form or stored on computer discs, hard drives, magnetic tape or digital or any other electronic medium.
	Any Data made available under this Agreement as defined on the cover sheet, to the exclusion of any communication Data automatically

Wizard Ansatz für
das Erstellen des
DSA's

DSA Template
vom IGE

Limitierungen des Wizard Ansatzes

- Limitierte Funktionalität: Unterstützt “nur” beim Erstellen des DSA’s. Was ist wenn ich bereits ein eigenes Template habe?
- Keine Möglichkeit mein eigenes Template zu validieren oder mit “best practice” vergleichen zu können
- Relativ viel manueller Aufwand beim Hinterlegen eines neuen Templates (Übersetzung in “enforceable policies”)
- Der Wizard nimmt keinen Bezug zu den Daten (resp. Metadaten)

→ Verwendung von GenAI ermöglicht neue Funktionalitäten und erhöht gleichzeitig “user experience”

CoSteward

Funktionalitäten

Der CoSteward ist eine GenAI basiertes Tool um Data Stewards beim Umgang mit Data Sharing Agreements zu unterstützen

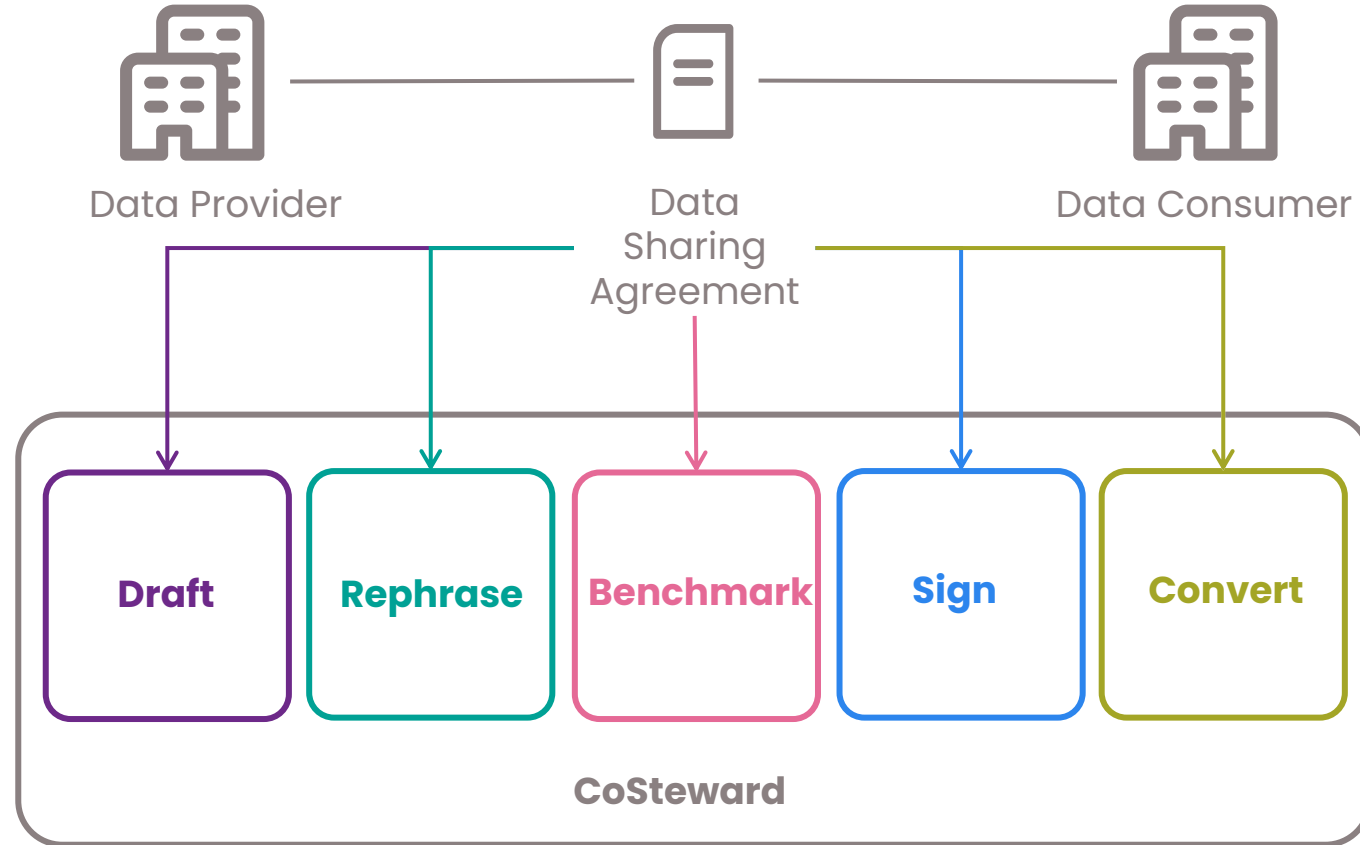
Draft – Buchstäblich auf einer leeren Seite starten

Rephrase – Bereits existierende DSA's validieren und anpassen

Benchmark – Bestehendes DSA mit "best practice" zu vergleichen

Sign – DSA elektronisch zu signieren

Convert – "enforceable policies" automatisch generieren lassen (i.e. ODRL)



CoSteward

Vorgehen

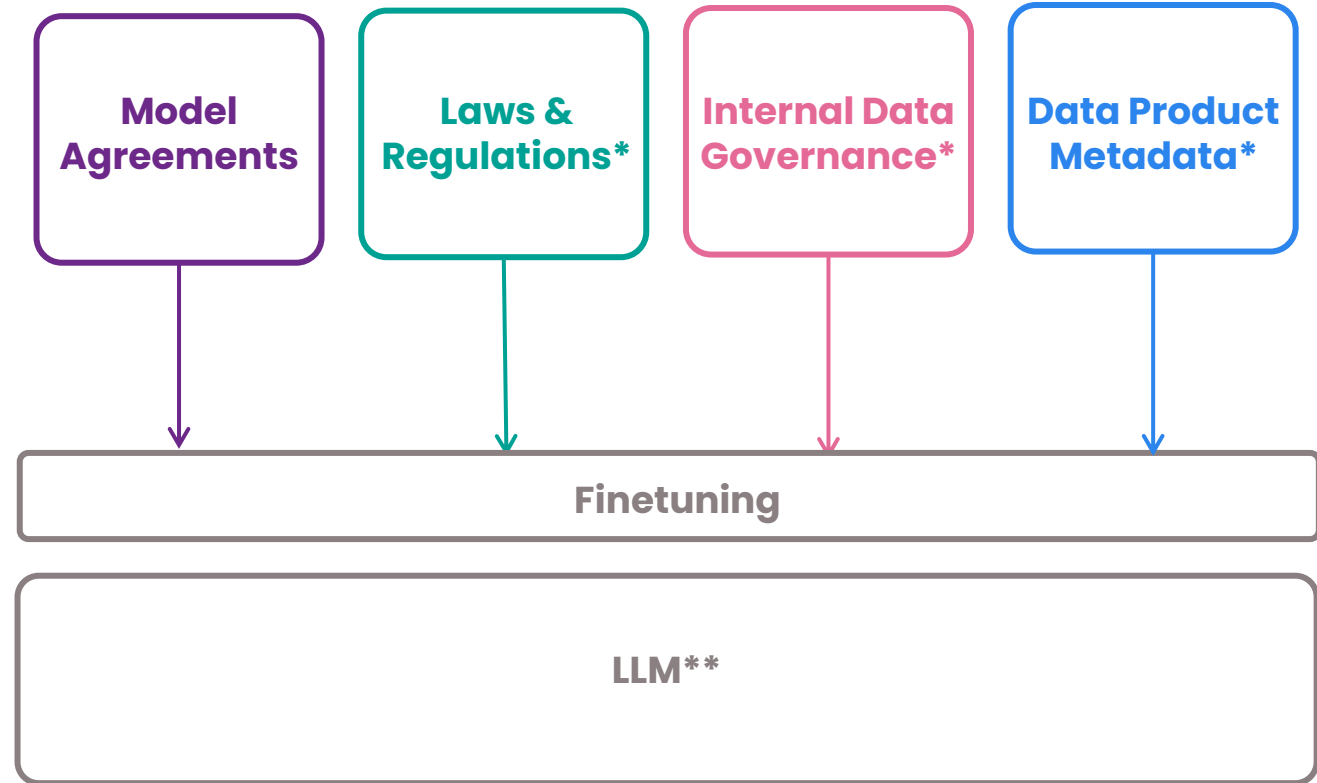
“Finetuning” des Large Language Model (LLM) basierend auf:

Model Agreements – Kuratierte Muster DSA’s (z.B. vom IGE) in verschiedenen Sprachen

Laws & Regulations* – Relevante Gesetze und Regulation für das jeweilige Datenaustausch-Szenario

Internal Data Governance* – Berücksichtigung der organisationsinternen Daten-Governance

Data Product Metadata* – Einbezug der Daten (resp. Metadaten)



* noch nicht vollständig implementiert

** Auswahl des LLM’s (i.e. GPT, Mistral AI, Silo AI)

CoSteward

Anwendung

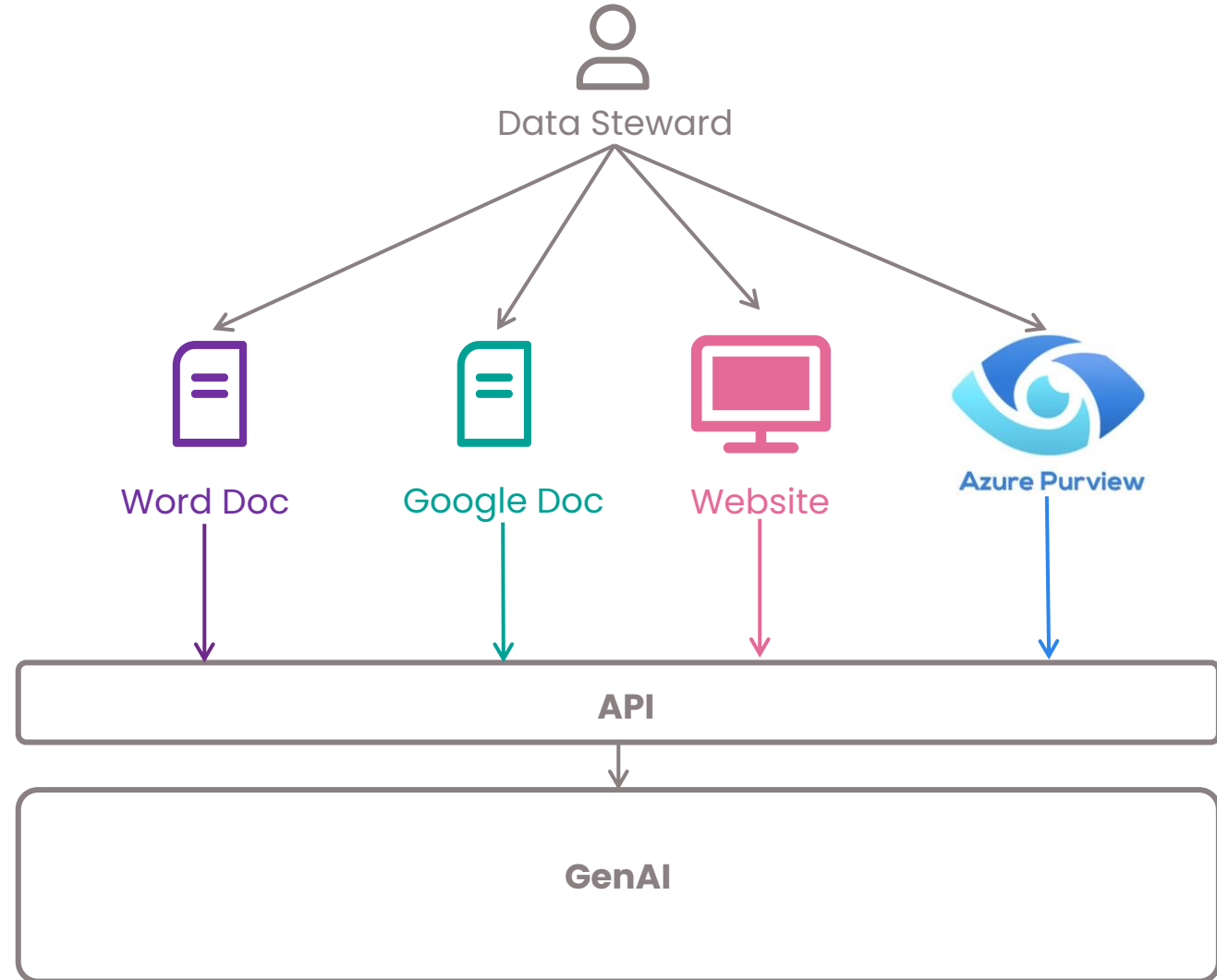
Der CoSteward ist nahtlos in die bekannten Applikationen integriert:

Word Doc – Ein Word Add-in verbindet den User mit dem GenAI Backend

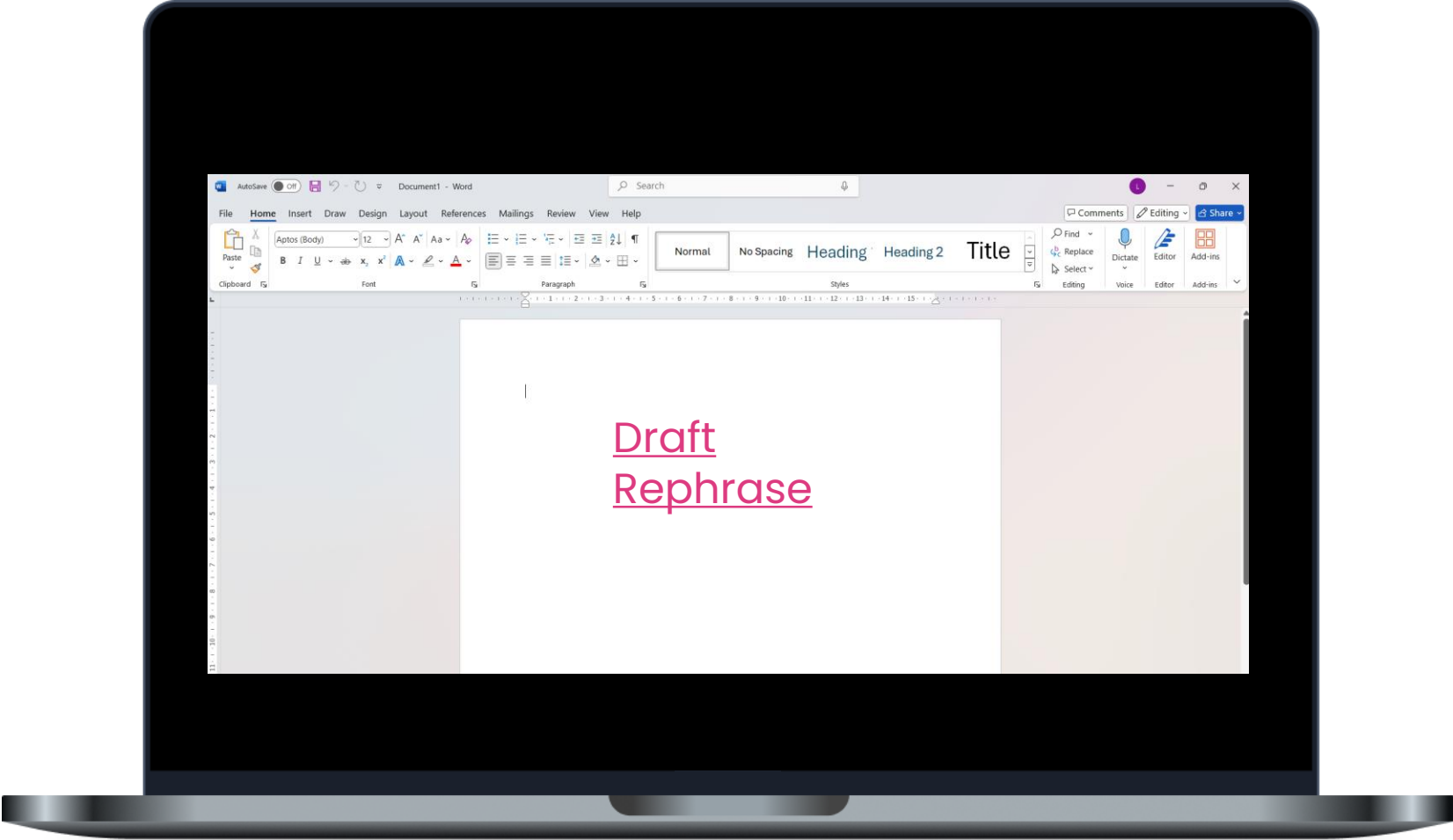
Google Doc* – Ein Google Doc Add-on verbindet den User mit dem GenAI Backend

Website* – Eine Website integriert das GenAI Backend via APIs

Azure Purview* – Integration in bestehende Data Governance Lösungen



CoSteward Demo



Swiss Data Space Forum

30 October 2024

Swiss Data Alliance

9:00 - 17:00 | Campus HSLU Rotkreuz

Kontakt Daten

Lukas Hohl

lukas.hohl@beyondcivic.com

Wilhelmshöhe 1, 6003 Lucerne

beyondcivic.com

Appendix