Wissenschaftliche Standardisierung und Data Provenance

Die Energiewirtschaft zählt zu den kritischen Infrastrukturen (KRITIS). Mit zunehmender Dezentralisierung stoßen traditionelle, silobasierte Stammdaten-Systeme an ihre Grenzen. Die Einführung von Künstlicher Intelligenz – insbesondere Agentic AI – zur automatisierten Konsistenzprüfung von Anlagenstammdaten erfordert ein neues Maß an Auditierbarkeit und Interoperabilität. Die Lösung für das sogenannte Vendor-Lock-in-Problem proprietärer Software-Silos liegt in der konsequenten Adaption wissenschaftlicher Datenstandards.

Cernion, als Referenzimplementierung für das Agentic Asset Master Data Management (a²mdm), löst diese Herausforderung architektonisch durch den **FAIR Data Layer** (eingeführt in v0.11 und finalisiert in v0.12).

Dieses Kapitel beleuchtet die technische Implementierung der FAIR-Prinzipien (Findable, Accessible, Interoperable, Reusable), die Integration der Open Energy Ontology (OEO) und die Sicherstellung von Data Provenance gemäß den Vorgaben des EU AI Acts für Hochrisiko-Systeme.

---

1. Vermeidung von Vendor-Lock-ins durch die Open Energy Ontology (OEO)

Ein zentrales Problem der Datenintegration bei Verteilnetzbetreibern ist das Fehlen eines gemeinsamen, maschinenlesbaren Vokabulars. Cernion begegnet diesem Problem durch eine native, statische Mapping-Schicht auf die **Open Energy Ontology (OEO)**, dem Standard-Vokabular der Nationalen Forschungsdateninfrastruktur (NFDI4Energy).

Technische Umsetzung (`oeo-mappings.js` & `sync-oeo.js`):
Cernion verzichtet bewusst auf einen ressourcenintensiven, vollumfänglichen Knowledge Graph (SPARQL-Endpoint) zur Laufzeit. Stattdessen zieht ein dediziertes Entwickler-Skript (`npm run sync:oeo`) die aktuellste OEO-ETD-CSV von GitHub und kompiliert daraus ein hochperformantes, statisches Lookup-Modul (`src/oeo-mappings.js`).

Dieses Modul umfasst ca. 150 kuratierte Mappings, die präzise auf die Anforderungen im Assetmanagement (z.B. Marktstammdatenregister / MaStR) zugeschnitten sind. Beispielsweise wird das MaStR-Feld `Lage` bei Windkraftanlagen nicht einfach als Text verarbeitet, sondern semantisch

differenziert in `wind-onshore` (OEO_00000446) und `wind-offshore`.

Semantische Anreicherung der KI (Classifier Enrichment):
Ein besonderer Architektureingriff liegt in der Nutzung der deutschen OEO-Labels (`labelDe`, z.B. „Umspannwerk“ oder „Solaranlage“). Diese Begriffe werden automatisiert in den Keyword-Pool der Cernion-Klassifizierungs-KI (Classifier) überführt. Dies boostet die heuristische Erkennungsrate bei unstrukturierten, deutschsprachigen Daten-Uploads von Netzbetreibern erheblich, ohne den eigentlichen Scoring-Algorithmus anpassen zu müssen.

OpenAPI-Annotationen:
Alle REST-Endpunkte (über 45 Actions in 7 Domänen wie `grid-operations` oder `energy-market`) sind via OpenAPI mit einer `x-oeo-class`-Erweiterung annotiert. Zusätzlich wird ein JSON-LD Endpunkt (`GET /api/datapoints/oeo-context`) bereitgestellt, der das Mapping von internen Cernion-Feldern auf OEO-IRIs als `@context`-Dokument ausliefert.

---

2. Data Provenance & KRITIS-Compliance (Der EU AI Act in der Architektur)

Systeme, die als Sicherheitskomponenten in der Verwaltung kritischer Infrastrukturen eingesetzt werden, unterliegen künftig den strengen Vorgaben des EU AI Acts für Hochrisiko-KI-Systeme (Art. 10, 12, 15). Die Kernanforderung lautet: **Data Provenance**. Es muss jederzeit kryptografisch nachweisbar sein, auf welcher Datengrundlage (Quellen, Lizenzen, Zeitstempel) ein KI-Agent eine Entscheidung (z.B. eine automatisierte Stammdaten-Korrektur) getroffen hat, um “KI-Halluzinationen” auszuschließen.

Die In-Memory-Edge-Architektur (PouchDB):
Cernion garantiert KRITIS-Compliance durch eine strikte In-Memory-Policy am Edge. Sensible Netztopologie- oder Assetdaten (z.B. im `Datapoint Layer`) verbleiben in der lokalen PouchDB-Instanz des Betreibers und werden nicht unmaskiert an externe Cloud-LLMs gesendet. Die Agenten operieren primär auf Basis der anonymisierten OEO-IRIs.

Der OEMetadata v2.0 Exporter (`oemetadata.service.js`):
Das Herzstück der Provenance-Sicherung ist der OEMetadata-Exporter (eingeführt in v0.11, strikt validiert in v0.12). Jeder verwaltete Cernion-Datenpunkt erzeugt auf Anfrage ein JSON, das exakt dem offiziellen `schema.json` v2.0 des OEMetadata-Projekts entspricht.

Der Exporter extrahiert dynamisch:

- **Temporal & Spatial Data:** Zeitreihen-Begrenzungen und Georeferenzierungen (Bounding Boxen) basierend auf den abgerufenen Datensätzen (z.B. das Netzgebiet „Ludwigshafen“).
- **Sources & Licenses:** Ein Cernion-Agent dokumentiert automatisch in den Metadaten, welche Microservices er abgefragt hat (`sources`). Das System mappt diese auf die korrekten Quell-Lizenzen (z.B. `DL-DE/BY-2.0` für das MaStR oder `CC-BY-4.0` für ENTSO-E).

- **Kryptografischer Hash:** Jeder Export beinhaltet einen kryptografischen Hash (z.B. SHA-256) der rohen Quelldaten zum Zeitpunkt der Agenten-Analyse, was die Auditierbarkeit für den Netzbetreiber rechtssicher macht.

Um sicherzustellen, dass keine invaliden Metadaten ausgeliefert werden, lädt das System während des Builds (`npm run sync:oemetadata`) das offizielle JSON-Schema von der Open Energy Platform herunter. Jede REST-Anfrage (`GET /api/datapoints/:name/oemetadata?validate=true`) wird zur Laufzeit via `ajv` validiert.

---

3. Orchestrierung über Systemgrenzen: Der OEP-Connector

Um KI-Agenten im Assetmanagement wirklich proaktiv handeln zu lassen, müssen sie die lokale Sicht (das eigene Netzgebiet) mit globalen Entwicklungen abgleichen können. Cernion v0.12 führt hierfür einen dedizierten, Read-Only-Connector (`oep.service.js`) zur **Open Energy Platform (OEP)** ein.

Dieser Microservice ermöglicht es dem Agent-Planner, komplexe Fragestellungen in mehrstufige Pläne (Steps) zu zerlegen.
Ein Prompt wie: *“Vergleiche die installierte PV-Leistung bei den TWL Netzen mit dem Netzentwicklungsplan-Szenario (NEP) 2035”* triggert nun folgende automatisierte Kette:

1. Der Agent fragt den lokalen Cernion-Service (`assets.solar`) für das MaStR-Netzgebiet ab.
2. Der Agent nutzt den OEP-Connector (`oep.query`), um das öffentliche NEP-Szenario direkt aus der Datenbank der nationalen Forschungsinfrastruktur zu ziehen.
3. Das LLM (als Reasoning-Layer) interpretiert das Delta lokal und speichert das Ergebnis als validierten Datenpunkt zurück in die PouchDB.

Um die Latenz bei Agenten-Entscheidungen gering zu halten, werden die OEP-Tabellenmetadaten in Cernion im RAM gecacht (TTL von 24h), während die eigentlichen Tabellendaten aus Datenschutzgründen lediglich durch den Speicher fließen (Pass-through).

---

4. Datapoint Scheduling: Kontinuierliche Datenhygiene

Für den operativen Betrieb eines Stadtwerks reicht ein manueller Aktualisierungs-Trigger nicht aus. KI-Agenten können nur dann belastbare Konsistenzprüfungen durchführen, wenn das Daten-Fundament (“Datapoint Layer”) stets aktuell ist.

Mit v0.12 implementiert Cernion ein lokales Interval-Scheduling (`strategy: "interval"`). Der `datapoint.service.js` prüft kontinuierlich (z.B. in einem 60-Sekunden-Tick), ob Datenpunkte veraltet sind. So kann beispielsweise eine “Tägliche Portfolio-Inventur (PV, Wind, Speicher)” konfiguriert werden (`intervalMinutes: 1440`). Ist ein Datenpunkt veraltet, markiert der interne Health-Endpoint diesen als `stale` und triggert autonom einen Refresh.

(Dieser Mechanismus ist essentiell für die Vorbereitung auf Cernion v0.14+, in dem autonome Validierungsagenten kontinuierlich und unbeaufsichtigt auf der Datapoint-Schicht nach Fehlern in der Marktkommunikation oder dem MaStR suchen werden).

Investitionssicherheit durch FAIR Data

Mit der Architektur des FAIR Data Layers positioniert sich Cernion als Referenzimplementierung für das Agentic Asset-MDM. Kein anderer kommerzieller Energiedaten-Agent erzeugt aktuell vollautomatisch OEMetadata-konforme Exporte inkl. OEO-Annotationen und Data-Provenance-Hashes. Für Stadtwerke bedeutet dies: Investitionssicherheit (kein Vendor-Lock-in), lückenlose Nachvollziehbarkeit für Audits (EU AI Act) und sofortige Interoperabilität mit der nationalen Energieforschungslandschaft.

Revision #1
Created 29 March 2026 14:27:06 by Thorsten Zoerner
Updated 29 March 2026 14:28:19 by Thorsten Zoerner