

Die KI-Datenplattform Baden-Württemberg

Daten- und Modellinventare intelligent vernetzen

Ergebnisbericht eines Teilvorhabens der KI-Allianz Baden-Württemberg eG, gefördert durch das Ministerium für Wirtschaft, Handwerk und Tourismus Baden-Württemberg. Ausgabe mit Schwerpunkt auf den Beiträgen des Fraunhofer IPA.

Ausgangslage und Zielsetzung

Viele Unternehmen, insbesondere kleine und mittlere Unternehmen (KMU), aber auch Kommunen, Behörden und Forschungseinrichtungen verfügen über wertvolle Datenbestände und KI-Modelle (KI-Assets), die oft nicht auffindbar oder nicht nutzbar sind. Die Vision des Projekts KI-Datenplattform (KI-DP) war eine technische Infrastruktur, die diese verborgenen Schätze zugänglich macht, den Anschluss an europäische Datenräume ermöglicht und die Wertschöpfung durch Künstliche Intelligenz (KI) fördert.

Das Projekt lief vom 20.07.2023 bis zum 30.06.2026, eingebettet in die KI-Allianz Baden-Württemberg eG und gefördert mit einem Gesamtvolumen von 5,8 Mio. €. Im Fokus stand eine Open-Source-Softwarelösung, mit der Unternehmen, Wissenschaft und Verwaltung ihre KI-Assets inventarisieren, teilen und verwerten können: unter Wahrung der digitalen Souveränität, unter Beachtung ethischer und rechtlicher Anforderungen und anschlussfähig an Standards und Initiativen wie IDSA, Gaia-X, Manufacturing-X und den European Health Data Space.

Die Drei-Ebenen-Architektur

Die Lösung ist modular aufgebaut und unterstützt drei Stufen der Wertschöpfung in Datenökosystemen:

1. **Level 1 (KI-Asset Angebote):** Die Ebene der Datenerzeuger. Onboarding-Werkzeuge überführen Daten aus heterogenen Quellen (Maschinensteuerungen, Sensorik, Excel) qualitätsgesichert in einen standardisierten Metadatenstrom. Die Originaldaten bleiben beim Eigentümer.
2. **Level 2 (KI-Asset Kataloge):** Das Herzstück der Plattform. Ein metadatengestützter Katalog auf Basis des Standards **DCAT-AP** macht Datensätze und KI-Modelle über intelligente Suchmechanismen auffindbar.
3. **Level 3 (Wissensplattform):** Die semantische Verknüpfung von Daten, Dokumenten und Expertenwissen über Retrieval Augmented Generation (RAG) schafft neue Beratungs- und Matchmaking-Dienste für die KI-Allianz und ihre Mitglieder.

Kernkomponenten der Plattform

Die technischen Ergebnisse wurden als Open-Source-Softwarepaket unter github.com/KI-Allianz veröffentlicht. Die Plattform basiert auf dem Piveau-Grundsystem und bietet folgende Schlüsselkomponenten:

- **Piveau.nextjs (Benutzeroberfläche):** Eine moderne Web-Oberfläche auf Basis von Next.js 15 mit intuitiver Suche, Datenvisualisierung und Favoritenverwaltung im Browser.
- **Semantische Suche:** Die Plattform versteht Konzepte statt nur Schlüsselwörter. Eine Anfrage nach „Bahn“ liefert auch Ergebnisse zu „Schienenverkehr“. Die Suche ist sprachübergreifend und kombiniert Vektorsuche mit klassischer Volltextsuche (Hybrid Search).
- **Datenqualitätsdienste:** Eine modulare Pipeline erkennt personenbezogene Daten (DSGVO-Compliance), ergänzt fehlende Werte, identifiziert Anomalien in tabellarischen Daten und bereitet sie für das Training von KI-Modellen vor.
- **Integrations-Hub:** Branchenspezifische Brückenwerkzeuge verbinden bestehende Datenräume mit dem Katalog: AAS2DCAT für die Industrie 4.0 (siehe Beitrag des Fraunhofer IPA), mds2hdcap für Gesundheitsdaten (NFDI4Health nach Health-DCAT-AP) und das FROST-Server DCAT-Plugin für IoT-Sensordaten aus Smart-City-Anwendungen.

Ethische und rechtliche Anforderungen

Ethische und rechtliche Aspekte wurden von Beginn an nach dem Prinzip „compliance by design“ integriert. Die Task Force Ethik & Recht (Universität Tübingen, FZI, Universitätsklinikum Freiburg) erarbeitete ein öffentlich verfügbares Whitepaper ([DOI: 10.6094/UNIFR/276339](https://doi.org/10.6094/UNIFR/276339)), das Plattformarchitektur, Governance-Strukturen und Metadatenanforderungen an Datensouveränität, Transparenz und Fairness ausrichtet. Ein rechtswissenschaftliches Kurzgutachten zu AI Act, Data Governance Act und Datenschutz ergänzt das Whitepaper. Die Plattform ist damit nicht nur technisch funktional, sondern auch regulatorisch robust.

Testbetrieb und Verfügbarkeit

Die Plattform wird seit Anfang 2026 unter ki-daten.hlrs.de auf der Infrastruktur des Höchstleistungsrechenzentrums Stuttgart (HLRS) öffentlich getestet. Der Katalog hält Metadaten zu über 7.000 Datensätzen vor, KI-Modelle aus Hugging Face wurden importiert. Der Testbetrieb läuft bis Ende 2026. Für den anschließenden operativen Betrieb, organisiert durch die KI-Allianz, wurden Geschäftsmodelloptionen erarbeitet.

Verwertung und Anwenderforum

Um den Praxistransfer zu sichern, wurde parallel zur Entwicklung ein Anwenderforum mit zehn Partnern etabliert. Unternehmen wie Averbis, EDI AG, PLEXPART, Matchory, NuCOS, raumobil, SSC-Services und Studio Fluffy testeten Nutzungsszenarien in den Bereichen medizinische KI, industrielle KI, Metall-3D-Druck, Mobilitätsdienste, Wissensmanagement und synthetische CAD-Daten.

Strategische Positionierung

Die Ergebnisse der KI-DP schaffen den Anschluss an weitere Initiativen:

- **Manufacturing-X:** Mit AAS2DCAT (Fraunhofer IPA) und der AAS-Infrastruktur FA³ST (Fraunhofer IOSB) erhalten KMU einen einfachen Einstieg in industrielle Datenräume.

- **Urbane Datenräume:** Die Stadt Freiburg ist als strategischer Partner über Level 1 angeschlossen. Eine Kooperation zur Übertragung auf die Stadt Karlsruhe ist vereinbart.
- **Gesundheit:** Das Universitätsklinikum Freiburg ermöglicht die Anschlussfähigkeit an MEDI:CUS, die cloud-basierte Medizininfrastruktur des Landes Baden-Württemberg.
- **Wissensplattform ask-bw:** Die auf den Hannover Messen 2025 und 2026 präsentierte Wissensplattform dient der KI-Allianz als Instrument für das regionale Matchmaking von KI-Experten und Anwendern.
- **Agentische KI-DP:** Unter Federführung des Fraunhofer IOSB wurde eine offene Multi-Agenten-Architektur konzipiert, die Recherche und Management von KI-Assets teilautomatisiert.

Der Beitrag des Fraunhofer IPA: KI in der Produktion

Das Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA leitete das Arbeitspaket „KI in der Produktion“ und brachte die Perspektive der produzierenden Industrie in die Daten- und Modellbeschreibung, die Datensuche und die Messe-Demonstratoren ein. Alle zugesagten Ergebnisse wurden erreicht.

- **Bedarf im Mittelstand:** In einer Umfrage des IPA unter produzierenden Unternehmen in Deutschland, Österreich und der Schweiz (2024, 6.159 kontaktierte Unternehmen, 137 Rückmeldungen) nannten die Teilnehmenden fehlendes Vertrauen, unzureichende Infrastruktur und regulatorische Unsicherheit als zentrale Hemmnisse für das Teilen von Produktionsdaten. Ergänzend sichtete das IPA rund 230 öffentlich zugängliche Produktionsdatensätze und evaluierte die Datenräume Catena-X, Manufacturing-X, Factory-X und IDS.
- **AAS2DCAT:** Das vom IPA entwickelte Open-Source-Werkzeug überführt Industrie-4.0-Verwaltungsschalen (Asset Administration Shell, AAS) aus FA³ST- oder BaSyx-Registries regelbasiert in den Metadatenstandard DCAT-AP und publiziert sie direkt in die KI-Datenplattform. Ein LLM-gestütztes Modul ergänzt Kategorien, Kurzttexte und Schlagworte, ein geführter Workflow mit Logging macht den Prozess nachvollziehbar. Die Datengeber behalten die Kontrolle über sensible Inhalte, zentral werden nur Metadaten geführt. AAS2DCAT ist seit November 2025 unter MIT-Lizenz auf GitHub verfügbar (github.com/KI-Allianz/aas2dcat). Ein einsatzfertiges Testpaket (Docker Compose für BaSyx und FA³ST, Builds für Windows und Linux, Benutzerhandbuch) erleichtert den Einstieg. Nutzertests im zweiten Halbjahr 2025 bestätigten die intuitive Bedienbarkeit und lieferten konkrete Verbesserungspunkte. Gemeinsam mit dem Fraunhofer IOSB wurde damit das Onboarding eines KMU aus dem Anwenderforum (NuCOS, Metall-3D-Druck) umgesetzt.
- **Datensouveränität und KI-Training:** Das IPA betrieb eine eigene Level-1-Instanz und erarbeitete mit dem HLRS eine sichere Rechen- und Trainingsumgebung, ergänzt um ein Konzept für datenraumkonforme Nutzungsrichtlinien (EDC/DCAT-AP-Policy), ein modulares Vertragswerk für Datennutzung und KI-Training sowie einen Geschäftsmodell-Blueprint „Industrial Data Monetization“.

- **Synthetische Produktionsdaten:** Am Use Case Batterieproduktion entstanden rund 800 synthetische Bilddatensätze (Deformationen, Kratzer, Schüttanordnungen) mit DCAT-AP-Beschreibung, die fehlende Realdaten für Entwicklung und Prototyping ersetzen.

Ausblick: Das IPA entwickelt AAS2DCAT als „Tool-as-a-Service“ für das niederschwellige Onboarding von KMU weiter, ausgerichtet an Manufacturing-X, Factory-X und Catena-X und getragen über die Transferkanäle der Fraunhofer-Gesellschaft sowie Kooperationen wie HammerHAI und Arena 2036. Darüber hinaus verfolgt das IPA die Konzepte eines „Asset Interface Hub“ für den einheitlichen Zugriff auf digitale Zwillinge und von „Asset Retrieval Agents“ auf Basis generativer KI.

Fazit

Das Projekt KI-Datenplattform hat eine skalierbare Software-Basis geschaffen, die verteilte KI-Assets sicher, standardisiert und auffindbar macht. Der Open-Source-Ansatz und die Anbindung an nationale und internationale Datenräume stärken die digitale Souveränität des Wirtschaftsstandorts Baden-Württemberg und senken die Einstiegshürden für KMU in ein KI-Ökosystem. Zusammen mit der KI-Allianz Baden-Württemberg eG laufen Gespräche mit interessierten Akteuren zum Übergang in einen operativen Betrieb.

Weiterführende Links:

- Projektseite: ki-allianz.de/ki-datenplattform
- GitHub: github.com/KI-Allianz
- AAS2DCAT: github.com/KI-Allianz/aas2dcat
- Testbetrieb: ki-daten.hlrs.de
- Magazinbeitrag zum Open-Source-Release: ki-allianz.de/opensource-veroeffentlichung-2025
- Whitepaper Ethik und Recht: [DOI: 10.6094/UNIFR/276339](https://doi.org/10.6094/UNIFR/276339)
- Kontakt Fraunhofer IPA: Jannik Schwab, jannik.schwab@ipa.fraunhofer.de

Unsere Partner der KI-Datenplattform:

