

## Ermittlung von Flächenversiegelung als komplexe Aufgabe (Projekt SEAL)

DR. PETER LORKOWSKI

NIEDERSACHSEN WASSER KOOPERATIONS- UND DIENSTLEISTUNGSGESELLSCHAFT MBH

### Hintergrund

Die Niedersachsen Wasser erfasst als Tochtergesellschaft des OOWV (Oldenburgisch-Ostfriesischer Wasserverband) den Versiegelungsgrad für diverse Gemeinden. Neben herkömmlichen Methoden wie GIS-Analyse und Luftbildinterpretation soll dabei auch zunehmend das Potential von Künstlicher Intelligenz (KI) bzw. Machine Learning (ML) genutzt werden. Die Vielzahl und Heterogenität der für diesen Prozess in Frage kommenden Datenquellen und Methoden stellt dabei eine besondere Herausforderung dar. Nach anfänglich im Rahmen eines Proof of Concept durchaus vielversprechenden Ergebnissen mit ML müssen diese nun auf einer breiteren Ebene und im Hinblick auf deren Eignung zur tatsächlichen Gebührenermittlung evaluiert werden [4]. Eine sorgsame Auswahl und Kombination von Daten und Methoden sowie eine realistische Einschätzung der jeweils erreichbaren Aussagekraft soll verwertbare Ergebnisse liefern.

Für die konsistente Organisation und somit den kontrollierten Ablauf der hier erforderlichen Prozesskette sind umfassende Modellabstraktionen und Architekturentscheidungen notwendig [6]. Nur auf diese Weise kann die Aufgabe in ihrer zwangsläufigen Komplexität überhaupt bewältigt werden. Anforderungen der Nachvollziehbarkeit, Reproduzierbarkeit, Nachweisbarkeit und Prozess-Dokumentation sind die Treiber der hier vorgestellten Systemarchitektur.

### Abstraktion von Daten und Prozessen

Die abflusswirksamen Flächen auf einem Grundstück als Grundlage zur Erhebung von Gebühren werden als Abfolge von Arbeitsschritten und Zwischenergebnissen ermittelt. Die in diesem Zusammenhang relevanten Daten und Prozesse lassen sich grundsätzlich vier Bereichen zuordnen, für die jeweils spezifische Eigenschaften gelten (siehe Abbildung 1):

- Der Bereich der externen Daten unterliegt nicht dem Einfluss der Systembetreiber und dient lediglich als Input-Quelle für alle weiteren Prozesse. Spezifikationen müssen somit übernommen werden. Wie in Abbildung 1 angedeutet, werden auch externe Daten durch einen Prozess generiert, dessen Ausgestaltung und Turnus (z. B. Bildflüge) wiederum meist nicht beeinflusst werden können.
- Interne Prozesse können gemäß der systemeigenen Spezifikationen konfiguriert, reproduziert, protokolliert und evaluiert werden. In diesem Bereich liegt die Gestaltungsfreiheit in Bezug auf Prozessoptimierung.
- Zwischendaten erfüllen als Kettenglieder innerhalb der gesamten Prozesskette eine zentrale Funktion. Sie enthalten den zu ihrer jeweiligen Erzeugung aufgewendeten Arbeitsaufwand (durch Mensch bzw. Maschine) als Mehrwert. Die Quantifizierung dieses Mehrwertes macht Aufwandsabschätzungen für andere, ähnliche Aufgaben möglich. Durch Aktualisierung von weiter vorne in der Prozesskette befindlichen Datenbeständen können die nachfolgenden Kettenglieder kaskadierend obsolet werden (siehe auch Abbildung 3).
- Schließlich stellen die kostenwirksamen Daten den eigentlichen Beweggrund zur Entwicklung des hier vorgestellten Systems dar. Diese Daten wirken sich direkt auf den Geschäftsprozess der Gebührenerhebung aus und müssen daher einem hohen Qualitätsanspruch gerecht werden. Sie sind jedoch für sich selbst genommen wertlos und unterliegen einer durch die Nachvollziehbarkeit ihrer Entstehung einzulösende Rechenschaftspflicht (Erklärbarkeit).

Ermittlung von Flächenversiegelung als komplexe Aufgabe (Projekt SEAL)

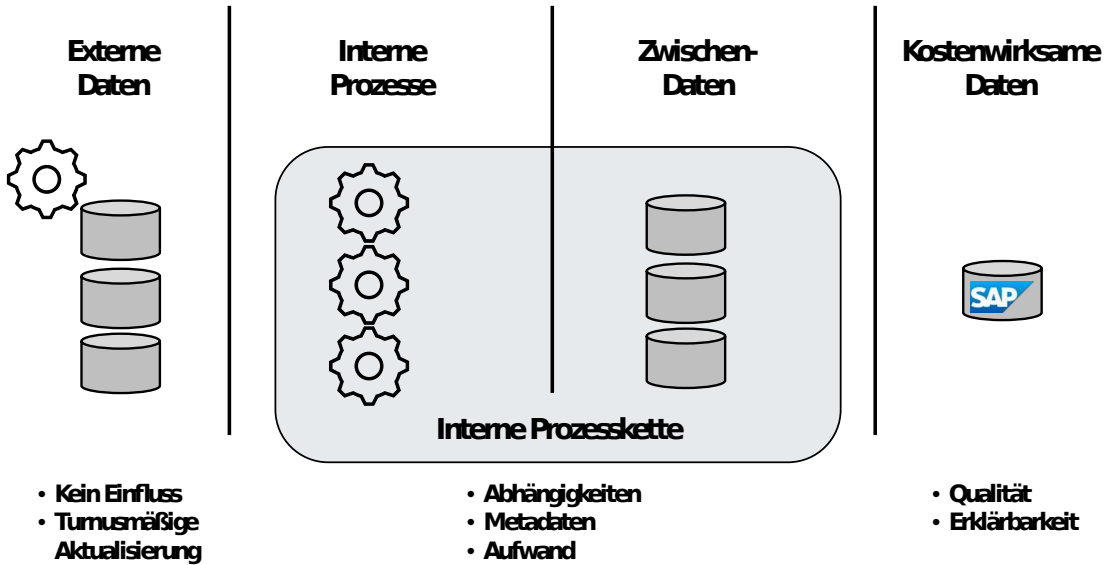


Abbildung 1: Kontexte von Daten und Prozessen nach Herkunft, Zweck und anderen Eigenschaften. Die interne Prozesskette (grauer Bereich) repräsentiert dabei eine Sequenz von Prozessschritten, die einheitlich beschreibbar und steuerbar sein sollen.

Die Ermittlung der kostenwirksamen Daten muss also stets als turnusmäßiger Aktualisierungsprozess statt als statischer Datenbestand gedacht werden [4]. Umso bedeutender wird deshalb die auf Automatisierung, Reproduzierbarkeit und Konfigurierbarkeit abzielende Beschreibung der jeweiligen Arbeitsschritte. Diese Metadaten müssen innerhalb des hier beschriebenen Systems einem definierten Schema folgen, um diesen Zweck erfüllen zu können. Dabei kann zwischen den Eigenschaften von Daten und den Eigenschaften von Prozessen unterschieden werden (siehe Abbildung 2). Während bei Daten eher beschreibende Eigenschaften wie raumzeitliches Bezugssystem, Struktur und Herkunft im Vordergrund stehen, werden bei einem Prozess diejenigen Daten als Metadaten gesehen, die der Methodenspezifikation, Steuerung und Konfiguration dienen.

Das KI-Modell (Layer von Neuronen-Gewichten [1]) kann in diesem Zusammenhang sowohl Daten als auch Steuerdaten zugeordnet werden. Es wird jedoch innerhalb der Prozesskette als Ergebnis eines aufwendigen Trainingsprozesses [1] mit Orthofotos und gelabelten bzw. Ground-Truth-Daten als Eingangsdaten erstellt. Obwohl dieses Modell dabei eher den Charakter einer Prozesssteuerung beim Inferencing statt von echten Eingabedaten hat, erscheint die Zuordnung zur Daten-Kategorie hier insgesamt sinnvoller (Abbildung 2). Auf diese Weise kann auch die Entstehungsgeschichte von KI-Modellen formal beschrieben werden.

## Ermittlung von Flächenversiegelung als komplexe Aufgabe (Projekt SEAL)

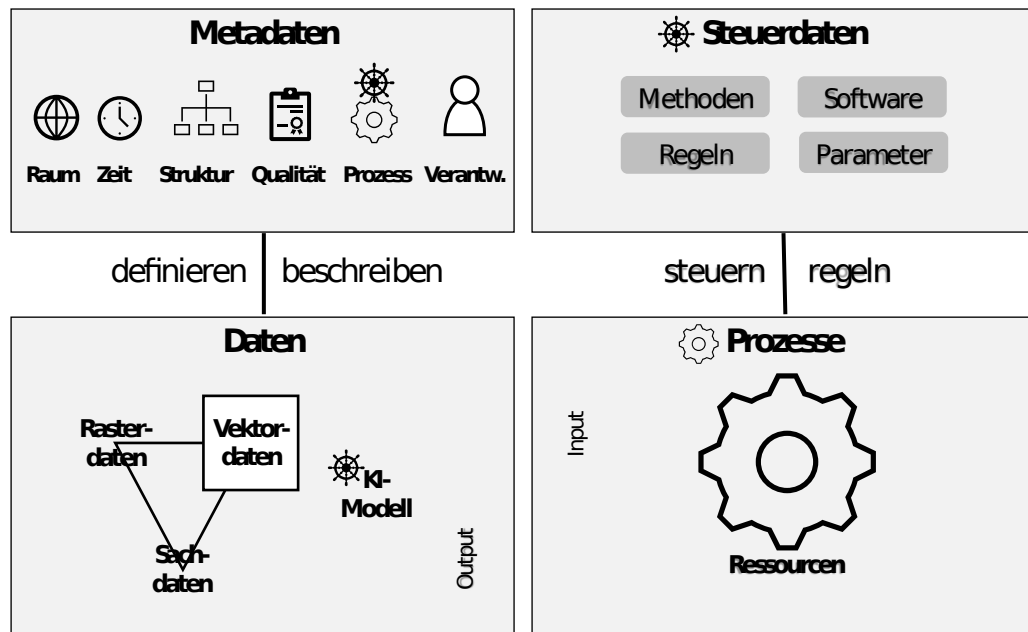


Abbildung 2: Beschreibung von Daten und Prozessen mittels Metadaten und Steuerdaten

Aufgrund der Input/Output-Beziehungen zwischen Eingabedaten, Prozessen und Ausgabedaten und der grundlegenden Abhängigkeitsstruktur ergibt sich für den Gesamtprozess eine Kaskade oder Daten-Pipeline (siehe Abbildung 3) mit einer schrittweisen Zunahme des informationellen Mehrwertes. Die Validierung der KI-Ergebnisse stellt innerhalb der Prozesskette die besondere Bedeutung einer iterativen Optimierung des KI-Prozesses zu [1].

Insgesamt dient die beschriebene Abstraktion innerhalb des Projektvorhabens folgenden Zwecken:

- Die Komplexität soll durch den so geschaffenen Ordnungsrahmen mental besser zu bewältigen sein [6].
- Eine effiziente Kommunikation mit Projektbeteiligten und Stakeholdern wird gefördert [5][6].
- Durch die Definition eines projektspezifischen Metadaten-Standard kann eine konsistente und qualitativ hochwertige Dokumentation gewährleistet werden, die den Nutzwert der Daten deutlich erhöhen kann.
- Die formale deklarative Beschreibung von Prozessschritten ist die Voraussetzung zur reproduzierbaren und konfigurierbaren Automation einer solchen Prozesskette.
- Das Erfassen des Aufwandes (Arbeitszeit von Personen, CPU-Zeit) eines Prozessschritts erlaubt eine Aufwandsschätzung bei Arbeitspaketen mit anderem Datenvolumen und erlaubt auch Aussagen über die Skalierbarkeit [2][3].
- Die strukturierte und standardisierte Beschreibung von Daten und Parametern ermöglicht eine bessere Kontrolle der Prozesskette [3].
- Arbeitsschritte und Prozessläufe können anforderungsgerecht dokumentiert und deren Zwischenergebnisse systematisch archiviert werden.

## Ermittlung von Flächenversiegelung als komplexe Aufgabe (Projekt SEAL)

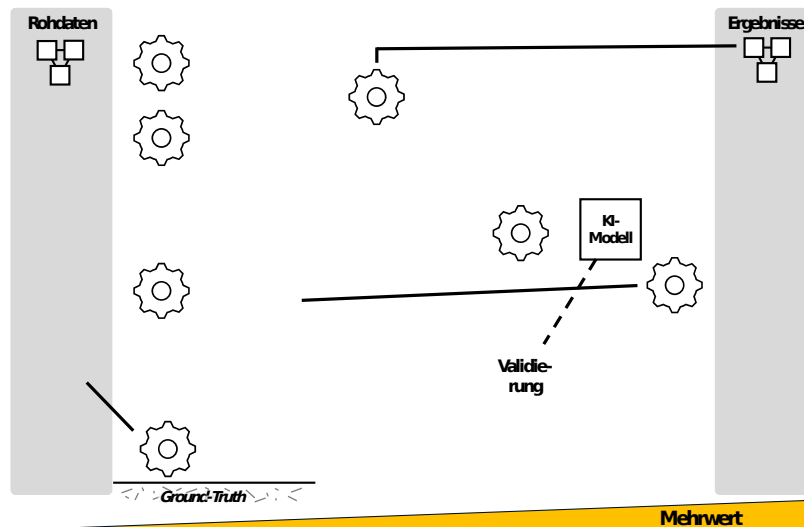


Abbildung 3: Prozesskette mit Daten-Elementen (Kästen) und Prozess-Elementen (Zahnräder). Die Validierung dient dabei der iterativen Verbesserung des ML-Modells.

### Umsetzungsstrategie

Das hier ausgeführte Beschreibungsschema der Prozesskette zur Ermittlung des Versiegelungsgrads setzt den Rahmen für einen systematischen und nachvollziehbaren Umgang mit der insgesamt sehr komplexen Datenverarbeitungsaufgabe. Eine von konkreten Produkten, Formaten, Programmiersprachen und Methoden abstrahierende Betrachtung minimiert Abhängigkeiten und erlaubt einen flexiblen Einsatz von Einzelkomponenten [5].

Als Datenquellen werden derzeit folgende genutzt bzw. sind vorgesehen:

- Orthofotos (z. T. TrueDOP, z. T. NIR-Kanal)
- DGM und DOM
- ALKIS-Daten (Tatsächliche Nutzung und Flurstücke)
- ATKIS-Daten
- Kanaldaten (Verbrauchsstellen-Positionen)
- Kundendaten aus SAP

Als Software kommen QGIS mit Geopackages und PostgreSQL / PostGIS zum Einsatz. Die Metadaten werden derzeit noch mit prototypisch MS Access verwaltet und sollen künftig in SQLite bzw. PostgreSQL geführt werden. Als Programmiersprache für GIS-Prozesse wird Python, für den ML-Bereich die Keras-Bibliothek verwendet. Über eine entsprechend gestaltete API-Definition soll der Austausch einzelner Komponenten grundsätzlich ohne größere Seiteneffekte ermöglicht werden [5][6].

Von der Nutzung spezifischer einschlägiger Tools und Standards zur Prozessautomatisierung und Metadatenverwaltung wurde zum jetzigen Zeitpunkt noch abgesehen. Die strategische Ausrichtung ist noch ausschließlich anforderungsgetrieben und auf die Vermeidung jeglicher technischer Abhängigkeiten ausgerichtet. Da aufgrund dieses Ansatzes mit Sicherheit zu erwarten ist, dass sich jetzt noch nicht bekannte Anforderungen ergeben werden, würde eine zu frühe Entscheidung zugunsten einer bestehenden Lösung u. U. mit hohen Anpassungsaufwänden verbunden, falls diese Lösung die Anforderung nicht erfüllt. Dies widerspräche der grundsätzlich auf Agilität ausgerichteten Projektstrategie [5].

Die Auswahl und Steuerung der gewählten Verarbeitungsmethoden von Datenbank-, GIS- und KI-Prozessen stellt angesichts der Vielzahl von Produkten und deren jeweiligen Konfigurationsmöglichkeiten bereits eine große fachliche Herausforderung dar. Diese Auswahl und Steuerung nicht jeweils über ein

## **Ermittlung von Flächenversiegelung als komplexe Aufgabe (Projekt SEAL)**

GUI-Menü auszuwählen, sondern aus einer expliziten persistierten Konfiguration aus einer Datenbank heraus durchzuführen stellt ein Kernelement der Prozessautomatisierung dar. Für die konsistente Verwaltung, aber auch für die experimentelle Permutation von umfangreichen und ggf. hierarchisch organisierten Prozess-Parametern wird ab einem gewissen Komplexitätsgrad eine rekursiv innerhalb eines Composite-Patterns [5] organisierte Speicherstruktur zum Einsatz kommen müssen [3]. Hier wird die besondere Herausforderung darin bestehen, auch bei Nutzung verschiedenster Softwarepakete und Berechnungsvarianten den jeweils verwendeten Parametersatz sowohl gegenüber Mensch als auch Maschine konsistent zu gestalten.

### **Zusammenfassung und Fazit**

Die Ermittlung versiegelter Oberflächen stellt eine Herausforderung in Bezug auf das Daten- und Prozessmanagement unter Verwendung unterschiedlichster Datenquellen und -formate, Software-Produkte und Konfigurationsmöglichkeiten dar. Das hier vorgestellte Rahmenwerk spiegelt die wesentlichen strategischen Entscheidungen zugunsten einer generischen und flexiblen Systemarchitektur für diese Aufgabe wider. Spezifische Beschreibungsschemata für Daten und Prozesse sowie deren gegenseitige Abhängigkeiten fördern das Systemverständnis und bilden ein beliebig konfigurierbares Kontrollinstrument zur Spezifikation und Reproduzierbarkeit auch komplexer Prozessketten. Die Implementierung der hier vorgestellten Komponenten verfolgt eine auf minimale Abhängigkeiten abzielende Strategie. Eine von bestehenden Lösungen abstrahierende Grundstruktur soll den Austausch von Komponenten erleichtern und als grundlegende Architekturmerkmale sowohl Stabilität als auch Flexibilität gleichermaßen bieten.

Kontakt zum Autor:

Dr. Peter Lorkowski  
Niedersachsen Wasser Kooperations- und Dienstleistungsgesellschaft mbH  
Georgstraße 4  
26919 Brake  
Fon: 0160/90566923  
E-Mail: lorkowski@niedersachsen-wasser.de

### **Literatur**

- [1] Kamusoko, Courage: Explainable Machine Learning for Geospatial Data Analysis, Boca Raton FL, 2025
- [2] Brinkhoff, Thomas: Geodatenbanksysteme in Theorie und Praxis, 4., überarbeitete und erweiterte Auflage, Berlin, 2022
- [3] Lorkowski, Peter: Monitoring Continuous Phenomena: Background, Methods and Solutions, Boca Raton FL, 2021
- [4] Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V. (DWA): Arbeitsblatt DWA-A 102-2/BWK-A 3-2, Grundsätze zur Bewirtschaftung und Behandlung von Regenwetterabflüssen zur Einleitung in Oberflächengewässer – Teil 2: Emissionsbezogene Bewertungen und Regelungen, Hennef, 2022
- [5] Evans, Eric: Domain Driven Design: Tackling Complexity in the Heart of Software, Boston, 2004
- [6] Taylor, R. N.; Medvidovic, Nenad; Dashofy, Eric: Software Architecture: Foundations, Theory and Practice, Danvers MA, 2010