

# Übertragung eines Vorgehensmodells zur KI-Integration von der Industrie auf Umweltinformationssysteme

Désirée Hilbring,<sup>1</sup> Julius Pfrommer<sup>2</sup>

**Abstract:** Maschinelles Lernen stößt in der Umwelt-Domäne auf großes Interesse. Allerdings ist der Einsatz von entsprechenden Algorithmen in Umweltinformationssystemen (UIS) bisher nicht weit verbreitet. Eine offene Frage ist zum Beispiel wie von Umweltbehörden bestehende Informationssysteme mit seit Jahrzehnten aufgebauten und gepflegten Umweltdatenbanken in einem hierarchischen föderalem System weiterentwickelt und für den Einsatz von neuen Technologien fit gemacht werden können. Hierbei sind nicht nur technische Aspekte von Interesse, sondern vor allem auch die Entwicklung geeigneter Prozesse in der Zusammenarbeit verschiedener Behörden. Für die Industrie wurde mit dem ML4P Vorgehensmodell ein toolgestütztes Verfahren für die Einführung von maschinellem Lernen in der Produktion entwickelt. Dieser Artikel untersucht dieses Vorgehensmodell im Hinblick der Übertragbarkeit des Modells auf die Einführung von maschinellem Lernen in von Behörden betriebenen Umweltinformationssystemen.

**Keywords:** Umweltinformationssysteme; Maschinelles Lernen; Machine Learning for Production (ML4P); Vorgehensmodell

## 1 Einleitung

Die öffentlichen Verwaltungen haben die letzten Jahrzehnte genutzt, im Bereich des Umweltmonitorings für die Erfüllung ihrer täglichen Aufgaben Fachinformationssysteme und Umweltdatenbanken aufzubauen. Die Funktionen behördlicher Systeme orientieren sich dabei primär an der Erfüllung von gesetzlichen Pflichten. Ein typisches Beispiel ist die Umsetzung der Wasserrahmenrichtlinie (WRRL), welche eine integrierte Gewässerschutzpolitik in Europa über Staats- und Ländergrenzen hinweg zum Ziel hat [PE00]. Dafür sind komplexe fachspezifische Algorithmen, Visualisierungen und Plausibilisierungen von Fachdaten unumgänglich. Die zugehörigen Fachalgorithmen werden von Experten häufig in Gremien abgestimmt und in den Systemen der nachgeordneten Behörden durch Software-Entwickler umgesetzt.

KI-Anwendungen im Umweltbereich sind verglichen mit anderen Disziplinen seltener. In den letzten Jahren werden jedoch verstärkt Methoden im Bereich Machine Learning

---

<sup>1</sup> Fraunhofer IOSB, Abteilung Informationsmanagement und Leittechnik, Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland, [desiree.hilbring@iosb.fraunhofer.de](mailto:desiree.hilbring@iosb.fraunhofer.de)

<sup>2</sup> Fraunhofer IOSB, Abteilung Informationsmanagement und Leittechnik, Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland, [julius.pfrommer@iosb.fraunhofer.de](mailto:julius.pfrommer@iosb.fraunhofer.de)

entwickelt. Eine Studie des Umweltbundesamtes „Künstliche Intelligenz im Umweltbereich“ untersucht die Forschungsaktivität aus dem KI-Bereich im Sinne der Nachhaltigkeit. Gefundene Themengebiete sind die Verbesserung des Recycling in der Abfallwirtschaft, der Einsatz von Verfahren der Mustererkennung bei der Erdbeobachtung und Naturkatastrophen, die Erhöhung von Energie- und Gebäudeeffizienz oder die Verbesserung von Produktionsverfahren zur Erhöhung der Nachhaltigkeit [TJ19].

Für die Integration von KI-Methoden in existierende produktive Umweltinformationssysteme ist neben der Entwicklung geeigneter Verfahren ein abgestimmtes Vorgehensmodell für die Anpassung bestehender Prozesse notwendig.

Die Anforderung existierende produktive Systeme mit neuen Methoden zu erweitern besteht auch in der Industrie. Deswegen wurde mit dem ML4P Vorgehensmodell ein toolgestütztes Verfahren für die Einführung von maschinellem Lernen in der Produktion entwickelt. Dieser Artikel analysiert die Prinzipien und Komponenten des Vorgehensmodells auf die Übertragbarkeit in Umweltinformationssysteme. Für die Diskussion der Abbildung werden beispielhaft Gewässerinformationssysteme herangezogen, welche in den Umweltverwaltungen der Bundesländer Baden-Württemberg, Bayern und Thüringen für die Umsetzung der Wasserrahmenrichtlinie eingesetzt werden [Us05].

Nachfolgend wird zunächst der als Beispiel herangezogene Anwendungsfall erläutert, ein Überblick über den Stand der Wissenschaft gegeben und anschließend das ML4P Vorgehensmodell vorgestellt, welches im Abschnitt „Abbildung des ML4P Vorgehensmodells auf Umweltinformationssysteme“ hinsichtlich der Aspekte Rollenkonzept, Phasenmodell, Maschine Learning Pipeline Diagramm, virtuelle Prozessakte und technischer Tools auf die Übertragbarkeit auf UIS untersucht wird. Der Artikel schließt mit Empfehlungen für nächste Schritte.

## **2 Anwendungsfall Umsetzung der WRRL in Gewässerinformationssystemen**

Die Abbildung des Vorgehensmodells wird in diesem Paper beispielhaft an folgendem Anwendungsfall diskutiert. Ziel der WRRL ist eine einheitliche Gewässerschutzpolitik in Europa zu erreichen, um die Gewässerqualität zu verbessern. Hierfür werden Vorgaben auf EU und Staatsebene definiert, welche in den einzelnen Staaten umgesetzt werden. Das bedeutet, dass der Ist-Zustand durch geeignetes Umweltmonitoring der Gewässer zu erfassen ist und der Soll-Zustand mit geeigneten Maßnahmen erreicht werden soll.

Für die föderale Struktur in Deutschland bedeutet dies, dass Bund und Länder ihr Handeln aufeinander abstimmen müssen. Dafür existieren länderübergreifende Fachgremien, welche die Vorgaben für die Umsetzung der Richtlinie definieren. Die aus der WRRL abgeleitete Erfassung und Verwaltung von Gewässerqualitätsdaten ist die Aufgabe der Länder und wird auf unterschiedliche Arten umgesetzt, z.B. durch die Entwicklung von geeigneten Umweltinformationssystemen. In den Bundesländern Baden-Württemberg, Bayern und

Thüringen existiert hierfür eine Entwicklungskooperation für die Realisierung von Gewässerinformationssystemen. Die resultierenden Fachanwendungen FIS GeQua, LIMNO und FIS Gewässer werden auf Basis der gemeinsamen Produktlinie WaterFrame® des Fraunhofer IOSB entwickelt [JS14]. Die Erfahrungen mit Weiterentwicklungen im Rahmen dieser real bestehenden Entwicklungskooperation dienen als Basis für die Abbildungsdiskussion des ML4P-Vorgehensmodells in diesem Artikel.

Die von der WRRL geforderten Daten und Analyseergebnisse werden zum Abschluss des Monitoring- und Analyseprozesses von allen Ländern in das bundeseinheitliche Tool WasserBLiCk gemeldet.

Als Hypothese für die Einführung von ML-Methoden wird als potentiell Beispiel die Idee der Einführung von Nitratprognosen in diesen Monitoringprozess diskutiert.

### 3 Stand der Wissenschaft

Das Umweltbundesamt hat mit seiner Studie „Künstliche Intelligenz im Umweltbereich“ Zukunftsperspektiven für den Einsatz von KI in der Domäne Umwelt untersucht [TJ19]. Die Studie stellt fest, dass wirtschaftliche Akteure Haupttreiber der Entwicklung von KI-Anwendungen sind. Dargestellt wird das Microsoft Programm „AI for earth“, oder eine Crunchbase-Analyse von Start-Up Unternehmen, die in der Umwelt-Domäne tätig sind und KI-Methoden einsetzen. Außerdem werden im Bereich der wissenschaftlichen Forschung die Möglichkeiten von KI erforscht. Im Abschnitt 4.2 „KI und (IT-)Infrastruktur für KI“ wird auf Voraussetzungen wie die Verfügbarkeit von Daten, die Nutzung leistungsstarker Rechenkapazitäten und das Vorhandensein sicherer Speicherarchitekturen eingegangen. Abschnitt 4.4 „KI als Sicherheits- und Warnsystem für den Umweltschutz“ stellt heraus, dass intelligente Systeme komplexe Zusammenhänge deuten und Muster erkennen können und damit als flexible Warnsysteme für den Umweltschutz dienen können.

Zieht man diese Aspekte in Betracht, ist es sinnvoll bestehende Datenbanken und Infrastrukturen bisher konventionell entwickelter Umweltinformationssysteme als Basis für die Einführung von ML-Methoden zu nutzen. Eine dafür potentiell geeignete Vorgehensweise aus der industriellen Produktion wird in den nächsten Abschnitten diskutiert. Alternative Vorgehensweisen für den Einsatz KI in behördlichen UIS sind in Veröffentlichungen bisher nicht zu finden.

Für die schriftliche Spezifikation von verteilten Software-Systemen existieren allerdings schon lange Vorgehensweisen für die Beschreibung von Software-Architekturen. Das Reference Model für Open Distributed Processing wurde in diversen Projekten (z.B. ORCHESTRA, SANY) für die Beschreibung von SOA-Architekturen angepasst, berücksichtigt aber nicht die speziellen Anforderungen der Integration von ML-Methoden [IS19] [Us07]. Unabhängig von der Vorgehensweise bei der Beschreibung gibt die Dissertation „Ein Rahmenwerk für die Architektur von Frühwarnsystemen“ einen Überblick über verschiedene Software-Architekturen von Umweltinformationssystemen [Mo17]. In der föderalen Struktur in Deutschland ist keine einheitliche technische Umsetzung von behördlichen UIS vorgeschrieben. Dies sollte ein Vorgehensmodell im Prozess berücksichtigen.

Existierende Vorgehensmodelle aus dem klassischen Data Mining, wie zum Beispiel CRISP-DM [WH00], sind anwendungsneutral und ohne den Hintergrund der industriellen Produktion entwickelt worden. Im Fraunhofer Leitprojekt *ML4P: Machine Learning for Production* wurde, gemeinschaftlich mit sechs beteiligten Fraunhofer Instituten, das toolgestützte *Fraunhofer Vorgehensmodell speziell für maschinelles Lernen in der Produktion* entwickelt. Das Ergebnis der Nutzung des Vorgehensmodells ist eine ML-basierte Anwendung im kontinuierlichen Betrieb in der Produktion. Das Vorgehensmodell kapselt Best Practices und ermöglicht die Skalierung auf große Teams durch entsprechende Planung, die Quantifizierung des Fortschritts und klare Schnittstellen zwischen Verantwortlichkeiten. Entsprechende Werkzeuge sorgen für die direkte Anwendbarkeit der Konzepte und eine hohe Geschwindigkeit in der Ausführung. In diesem Paper wird das Vorgehensmodell in seinen Grundzügen vorgestellt.

Eine zentrale Eigenschaft des ML4P Vorgehensmodell ist die Visualisierung des Umzusetzenden Prozesses in einem Maschine Learning Pipeline Diagramm, welches von Beginn an eine anschauliche zu verfeinernde und umzusetzende Kommunikationsbasis für alle Beteiligten bietet. Da dies durch die komplexe Verteilung von Zuständigkeiten in den Behörden ein wichtiger Aspekt bei der Einführung von ML-Methoden in Umweltinformationssystemen ist, wurde das ML4P Vorgehensmodell als Basis für die Untersuchung der Übertragbarkeit ausgewählt.

## 4 Beschreibung des ML4P Vorgehensmodells

ML4P ist ursprünglich für den Einsatz von maschinellem Lernen in der Produktionstechnik entwickelt worden, um über datengetriebene Prozessmodelle Verbesserungen einzubringen zu können. In der folgenden Beschreibung wird der Bezug zur Produktion auch ersichtlich sein. In der Folge findet die Übertragung auf Umweltinformationssysteme statt.

Das Vorgehensmodell ist in sechs Phasen mit klaren definierten Ergebnissen gegliedert. Für jede Phase ist eine Reihe von Phasenergebnissen definiert. Diese münden in einen Meilenstein, an dem die Phasenergebnisse begutachtet werden und eine Planung der nächsten Schritte vorgenommen wird [JP20].

Das Vorgehensmodell befindet sich in einem Spannungsfeld zwischen agilem und linearem Vorgehen. Viele ML-Projekte starten mit einer großen Unsicherheit bezüglich der Datenlage und der Eignung verschiedener Lösungsansätze. Diese Unsicherheit begünstigt ein agiles Vorgehen, bei dem im Projektverlauf laufend nachjustiert wird. Auf der anderen Seite sind Änderungen und Experimente an realen Produktionsanlagen kostenintensiv und setzen eine gute Planung voraus. Um dieses Spannungsfeld aufzulösen werden agiles und lineares Vorgehen kombiniert

1. Agiles Vorgehen innerhalb der Phasen

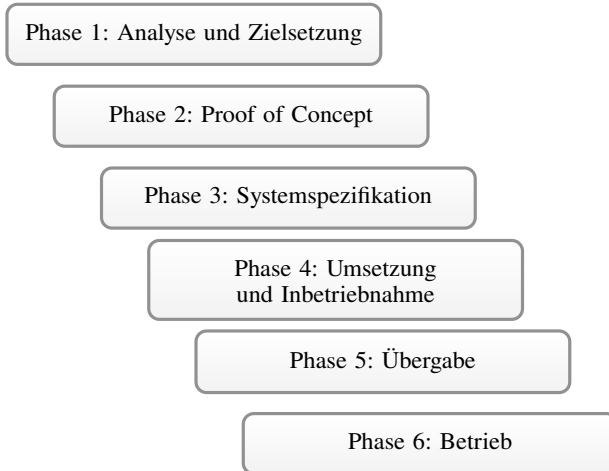


Abb. 1: Die sechs Phasen des ML4P Vorgehensmodells

## 2. Lineares Vorgehen über Phasen hinweg

Ein Zurückspringen auf vergangene Phasen sollte nur in Ausnahmefällen, wie etwa bei einer extern veranlassten Veränderung des Anwendungsbereichs und der Anwendungsziele, vorgenommen werden.

Das Ergebnis eines maschinellen Lernverfahrens hängt von der Verfügbarkeit qualitativ hochwertiger Datenbestände ab. In der industriellen Produktion ist eine nachträgliche Aufbereitung der Daten aber nur mit großem Aufwand machbar. Wenn zum Beispiel eine Zuordnung zwischen Datenströmen nicht möglich ist, kann eine nachträgliche Aufbereitung sogar unmöglich sein! Etwa wenn Messungen von Produktproben in externen Labors durchgeführt werden und deren Ergebnisse erst nach einigen Wochen zurück geführt werden. Nur bei einer guten Datenhaltung und mit Hilfe von geeigneten Datenmodellen ist bei zeitlich verzögerter Integration eine Zuordnung der Laboranalysen zu den Rahmenbedingungen des Probenahmezeitpunkt in der Produktion möglich. Ein entsprechendes Beispiel aus dem UIS Umfeld wäre die Zusammenführung der erfassten Umweltbedingungen zum Zeitpunkt der Probenahme mit dem zeitlich gestellten Rückfluss der analysierten Labor-Messwerten aus verschiedenen Proben des Probenahme.

Im Vorgehensmodell wird die Datenverarbeitung ganzheitlich im Sinne einer durchgängigen Verarbeitungskette betrachtet. Dafür wird bei der Durchführung der Phasen des Vorgehensmodells ein sogenanntes Machine Learning Pipeline Diagramm aufgestellt. Beispiele finden sich in [JP20]. Ein solches ML Pipeline Diagramm ist zunächst ein visuelles Hilfsmittel zur Unterstützung menschlicher Kommunikation und der Dokumentation des gemeinsamen Prozessverständnisses. Daneben sind ML Pipeline Diagramme ein gutes Mittel zur Kommunikation mit dem Management.

In kleinen Projekten können die Schritte von der Datenerfassung bis zur Modellbildung noch von Hand ausgeführt werden. Für den automatisierten kontinuierlichen Betrieb und die Verarbeitung großer Datenmengen ist dies jedoch nicht mehr möglich. Für den kontinuierlichen und wiederholbaren Betrieb wird die ML Pipeline auf konkrete Technologien abgebildet und umgesetzt.

Die Phasen des Vorgehensmodells beziehen sich auf die (Weiter-) Entwicklungen und den Betrieb einer Machine Learning Pipeline:

In Phase 1 wird die Ist-Situation und die Ziel-Situation für die ML Pipeline aufgestellt. Diese Aktivität ist ein wichtiger Schritt um ein gemeinsames Verständnis zwischen ML-Experten und Prozessexperten sicher zu stellen. Die Darstellung erfolgt in einem Machine Learning Pipeline Diagramm und ist dabei noch rein konzeptionell und unabhängig von konkreten Technologien. Hierbei ist es wichtig einerseits den aktuellen Ist-Zustand der in der Produktion genutzten Prozesse ohne ML-Methoden abzubilden und diesen von einem Entwurf des idealisierten Zielzustand zu unterscheiden.

In Phase 2 wird die ML Pipeline als Proof of Concept (PoC) verfeinert und validiert. Dafür findet eine partielle technologische Umsetzung statt. Diese ist aber in der Regel noch nicht für den dauerhaften Betrieb geeignet. Gängige Mittel zur Umsetzung des PoC umfassen die manuelle Ausleitung von Stichproben aus der Produktion und Datenabzüge von freistehenden (nicht vernetzten) Datenbanken der einzelnen Maschinen.

In Phase 3 wird die ML Pipeline in einer Systemspezifikation auf Technologien für den Dauerbetrieb abgebildet. Insbesondere Schnittstellen zwischen technischen Komponenten der ML Pipeline und Mensch-Maschine-Schnittstellen für die Interaktion mit dem Anlagenbediener sind im Detail zu betrachten. Die Systemspezifikation ermöglicht die Schätzung von Kosten einer Lösung für den dauerhaften Betrieb.

In Phase 4 wird die ML Pipeline als Bestandteil der Produktionsanlage für den kontinuierlichen Betrieb umgesetzt und in Betrieb genommen. Häufig wird durch Abhängigkeit zwischen Teillösungen eine gestaffelte Umsetzung gewählt. Zum Beispiel werden häufige strukturelle Anpassungen an Anlagen umgesetzt bevor die regelungstechnischen Anpassungen darauf aufgesetzt werden.

In Phase 5 wird die ML Pipeline an den Anlagenbetreiber übergeben. Dies ist ein sehr wichtiger Schritt um den Anlagenbetreiber zu befähigen das System langfristig zu betreiben. Dazu gehört auch die Schulung der Mitarbeiter in den entsprechenden Rollen.

Mit Beginn der Phase 6 ist die fertig implementierte ML Pipeline im Produktivsystem umgesetzt. Innerhalb der Phase 6 können Betrieb, sowie Wartung, Pflege und kontinuierliche Verbesserung der ML Pipeline durchgeführt werden. Dies ist notwendig insbesondere wenn der physische Produktionsprozess durch Verschleiß, Veränderung der Rahmenbedingungen oder strukturelle Umbauten im Laufe des Betriebs fortentwickelt wird.

**Durchgängige Artefakte** Neben den definierten Phasenergebnissen fußt das Vorgehensmodell auf zwei durchgängigen Artefakten — Dokumente und Datenstrukturen, die im Vorgehensmodell initial erstellt und in allen folgenden Phasen angewandt und fortgeführt werden. Sie repräsentieren den aktuellen Entwicklungsstand und sind zentral für das Wissensmanagement:

1. **Machine Learning Pipeline Diagramm:** Ein Machine Learning Pipeline Diagramm gibt eine Übersicht über die Datenerhebung und -verarbeitung, Modellbildung und Entscheidungsfindung in einem ML-unterstützten Produktionsprozess, siehe Abb. 3. Die verschiedenen Phasen erfordern ein Pipeline Diagramm in unterschiedlicher Granularität.
2. **Virtuelle Prozessakte:** Die virtuelle Prozessakte ist eine Informationsquelle, welche für das Projektteam relevantes Wissen über die Anlage bereithält und während der gesamten Projektzeit immer wieder adaptiert und aktualisiert wird.

**Rollenmodell** In den einzelnen Phasen werden unterschiedliche Kompetenzen und Disziplinen benötigt. In einem Rollenmodell sind die benötigten Disziplinen für die Phasen ausgeführt. Dabei kann eine Person auch mehrere Rollen ausfüllen. Das Fehlen einer Projekttrolle im Team ist ein starker Hinweis für zu erwartende Probleme und Verzögerungen im weiteren Projektverlauf.

**Technische Architektur und Werkzeugunterstützung** Das Vorgehensmodell zielt auf eine planbare und reproduzierbare Durchführbarkeit von ML-Projekten in der industriellen Produktion ab. Durch die Entwicklung von Best-Practices und strukturierter Vorgänge bietet es sich an, Projekte anhand des Vorgehensmodells durch einheitliche technische Werkzeuge zu unterstützen. Dafür wird eine Architektur vorgestellt, in der die Werkzeuge zu einer ganzheitlichen Lösung im Sinne einer Machine Learning Pipeline integriert werden.

## 5 Abbildung des ML4P Vorgehensmodells auf Umweltinformationssysteme

Dieser Abschnitt diskutiert eine mögliche Abbildung des ML4P Modells auf UIS im Hinblick auf essentielle Unterschiede. Das ML4P Vorgehensmodell beschreibt das Vorgehen für Produktionsanlagen. Umweltinformationssysteme hingegen bestehen nur zu einem kleinen Teil aus fachspezifischer Hardware (z.B. Umweltsensoren). Ihre zentralen Komponenten bestehen aus Software, deren verschiedene Funktionen für die Erfüllung von gesetzlichen Aufgaben dienen. Die Systeme sind an vielen Stellen mit Schnittstellen zu Softwaresystemen verschiedener Hersteller z.B. für Datenlieferungen oder spezielle Auswertungen ausgestattet.

## 5.1 Abbildung von Rollenkonzepten

Wichtig für einen funktionierenden Einführungsprozess ist die Definition wichtiger Projektrollen und die Zuordnung von entsprechenden Mitarbeitern. Genauso wie in der industriellen Produktion erfordert die Einführung von ML-Methoden in Umweltinformationssystemen die Zusammenarbeit verschiedener fachlicher Disziplinen. Die entsprechenden Mitarbeiter, arbeiten jedoch nicht zwangsweise in der gleichen Behörde. Eine Kommunikation über Behörden und Landesgrenzen hinweg ist notwendig. Dieser Abschnitt analysiert, inwieweit das ML4P Rollenkonzept auf existierende Strukturen abgebildet werden kann:

- Der Projektsponsor ist für die Finanzierung des Projektes zuständig. Im Umweltbereich sind dies häufig die Führungskräfte verschiedener Referate in den Landesbehörden.
- Prozessexperten sind die Fachexperten der Umweltinformationssysteme. Die Systeme sind komplex und realisieren eine Vielzahl von verschiedenen Monitoring und Auswertemöglichkeiten, um gesetzliche Vorgaben aus verschiedenen Fachbereichen erfüllen zu können. Der Prozessexperte ist also üblicherweise keine einzelne Person, sondern eine Gruppe von Angestellten der übergeordneten Landesbehörden, die Anforderungen aus gesetzlichen Grundlagen und anderen Anforderungen für die Umsetzung im UIS definiert.
- Die Prozessbediener nutzen die Funktionen des UIS im täglichen Betrieb, um die gesetzliche Vorgaben zu erfüllen oder Anfragen aus der Bevölkerung zu beantworten. Prozessbediener gibt es sowohl in übergeordneten als auch in nachgeordneten Landesbehörden. Die Personalressourcen in Behörden sind knapp, deswegen haben Prozessexperten üblicherweise eine Doppelrolle als Prozessbediener.
- Der Automatisierungsingenieur des ML4P Modells kann in nächster Näherung mit den System-Administratoren der Prozessexperten verglichen werden. Diese Personen bilden innerhalb der Gruppe der Prozessexperten eine Entscheidungsebene für den Betrieb der UIS und können die Konfiguration des UIS ggf. in Zusammenarbeit mit den Software-Entwicklern entsprechend anpassen.
- Im ML4P Vorgehensmodell nicht vorgesehen ist die Rolle des Softwareentwicklers. Diese Personen sind meist nicht in den Landesbehörden, sondern in Software-Firmen angestellt und für die Wartung und Weiterentwicklung der UIS zuständig. Je nach Aufgabe stehen sie in direktem Austausch mit den System-Administratoren der Prozessexperten und den Prozessexperten.
- Bei der Rolle des Verantwortlichen für IT-Sicherheit gibt es keine speziellen Unterschiede im Hinblick auf die Übertragung des Konzeptes auf UIS.
- Die ML-Experten sind eine neue Gruppe von Personen, die für die Integration von ML-Methoden in bestehende UIS in die eingespielten Rollen der Entwicklung von Umweltinformationssystemen integriert werden müssen. Notwendig für eine



erfolgreiche Integration ist eine Abstimmung mit Prozessexperten und Software-Entwicklern.

## 5.2 Abbildung des Phasenmodells

Für ein geeignetes Vorgehen bei der Einführung von ML-Methoden wurden im ML4P Vorgehensmodell sechs zentrale Phasen definiert, welche im Folgenden auf ihre Übertragbarkeit auf Umweltinformationssysteme untersucht werden.

Das in Phase 1 *Zieldefinition und Lösungsansatz* zu erstellende Konzept inklusive Beschreibung der vorhandenen Ist-Situation und einer gewünschten Soll-Situation ist für die Weiterentwicklung von Umweltinformationssystemen essentiell und kann übertragen werden. Es soll hierbei ein gemeinsames Verständnis von den mit dem UIS vertrauten fachlichen Prozessexperten, den Softwareentwicklern der betroffenen UIS Komponenten und den in dieser Phase das System kennennlernenden ML-Experten entstehen. Eine Beschreibung der Zieldefinition, eine initiale Versionen der Maschine Learning Pipeline Diagramm und der virtuellen Prozessakte bilden das Ergebnis dieser Phase.

Ausgehend von diesem konzeptionellen Verständnis sollen im ML4P Modell in einem agilen Prozess geeignete Lösungsansätze gefunden werden, wobei die zu testenden Lösungsansätze entwickelt, mit Testdaten erprobt und verfeinert werden, bis ein geeignetes Verfahren gefunden wird. Bei der konventionellen Weiterentwicklung von UIS werden bestimmte fachliche Lösungsansätze häufig durch übergeordnete Gremien festgelegt und für die Umsetzung in die Systeme vorgegeben. Für die im Anwendungsfall als Beispiel herangezogenen Gewässerinformationssysteme ist dies z.B. die Bund/Länder Arbeitsgemeinschaft Wasser - LAWA [Gr]. Ziel ist die einheitliche Umsetzung von EU- oder bundesweit vorgegebenen Monitoring oder Analyseaufgaben: z.B. die der Wasserrahmenrichtlinie. Die eigentliche Umsetzung wird im föderalen deutschen System an die nachgeordneten Behörden der Länder delegiert.

Bei der bisherigen Umsetzung der Vorgaben in den verschiedenen in den Ländern eingesetzten Gewässerinformationssystemen bestehen Freiheitsgrade, welche von Prozessexperten der Behörde, die das UIS betreibt, und Softwareentwicklern in einem iterativen und agilen Entwicklungsprozess mit von Prozessexperten definierten Testdaten über die Erzeugung von Piloten und Previews ausgestaltet werden können, bis ein geeignetes Umsetzungsergebnis erzielt wird. Bei der Einführung von ML-Methoden, welche unabhängig von anderen Systemen nur für ein bestimmtes Landes-UIS integriert werden sollen, kann das aus Entwicklungssicht bestehende Vorgehen als Phase 2 *Proof of Concept* des ML4P Modells problemlos übertragen werden.

Für die erfolgreiche länderübergreifende Einführung von ML-Methoden muss die bisherige länderübergreifende fachliche Abstimmung von vorgegebenen Lösungen angepasst werden, da ein iterativer Prozess zur Ermittlung eines geeigneten KI-Algorithmus notwendig ist. Hierfür werden Tests mit spezifischen UIS, wie z.B. einem Gewässerinformationssystem (FIS GeQua, LIMNO oder FIS Gewässer aus der Entwicklungskooperation) notwendig.

Daraus folgt, dass einerseits erfahrende Prozessexperten aus verschiedenen UIS bereits in Phase 1 *Zieldefinition und Lösungsansatz* als auch in die Phase 2 *Proof Of Concept* in die übergeordnete fachliche Arbeit der Gremien eingebunden werden sollten. Andererseits ist in diesen Phasen auch bereits die Beteiligung der ML-Experten notwendig. Einige spezifische UIS sollten als *Proof Of Concept* Kandidaten ausgewählt und in dieser Phase getestet werden. Die Besetzung der übergeordneten fachlichen Gremien muss also überprüft und erweitert werden. Das Ergebnis könnte eine ML Pipeline sein, die für die Umsetzung in den Behörden aller Länder in den späteren Phasen entsprechend verfeinert und angepasst wird.

Eine korrekte und konkrete Spezifikation von Schnittstellen ist für konventionelle Umweltinformationssysteme ebenfalls essentiell. Einerseits müssen die Schnittstellen zu Softwaretools verschiedener Hersteller für Softwareentwickler entsprechend dokumentiert werden, andererseits wird für die Bedienung neuer Funktionalitäten eine Anwenderdokumentation für Prozessbediener erstellt. Da die Umsetzung der Vorgaben in konkrete UIS zu den Aufgaben der Länder gehört, unterscheiden sich die länderspezifischen UIS und somit auch die Spezifikationen stark. Bei der Umsetzung des ML4P Vorgehensmodells schlagen wir vor, in dieser Phase die Entwicklung der Systemspezifikation an die zuständigen Prozessexperten und Softwareentwickler der Entwicklungskooperationen mit zugehörigen Behörden zu übergeben. Damit entsteht abhängig von der technischen Ausgestaltung eines UIS eine angepasste konkrete ML Pipeline pro technischer Umsetzung eines UIS. Für die Abbildung der UIS-Anforderungen an die jeweilige, zumeist serviceorientierte UIS-Infrastruktur einer Behörde oder eines Behördennetzwerks kann auf dedizierte Methoden wie z.B. SERVUS (Design Methodology for Information Systems based upon Service-oriented Architectures and the Modelling of Use Cases and Capabilities as Resources) zurückgegriffen werden [UB18]. Damit ist die Phase 3 *Systemspezifikation* übertragbar [JS14] [DHvdS19].

Ausgehend von der Systemspezifikation aus Phase 3 und den in Phase 2 entwickelten technischen Vorarbeiten kann in Phase 4 *Umsetzung und Inbetriebnahme* die eigentliche Implementierungsarbeit durch die Softwareentwickler und ML-Experten pro UIS stattfinden. Die dabei entstehenden neuen Auslieferungsversionen der Software werden von den fachlichen Prozessexperten der einzelnen Länder getestet und auftauchende Probleme an die Softwareentwickler und ML-Experten zurückgemeldet. In dieser Phase wird die neue Version in eine aus der Produktion gespiegelten Umgebung installiert.

Nach erfolgreichem Test erfolgt die Abnahme der Software durch die Prozessentwickler: Phase 5 *Übergabe*. Die Übernahme der Installation in den Produktionsbetrieb erfolgt dann üblicherweise durch die zuständigen IT-Referate oder Rechenzentren der Behörden der Prozessentwickler.

Damit ist die Übergabe vom Softwareentwickler und den ML-Experten an die Behörden der Prozessentwickler abgeschlossen und die Wartungsphase Phase 6 *Betrieb* beginnt. In dieser Phase können durch entsprechende Nachlieferungen Anpassungen und Fehlerbehebungen an einzelnen Softwarekomponenten umgesetzt werden.

Phasen	Hauptbeteiligte der Phase	Beteiligte Verwaltungseinheit der Phase
Phase 1: Analyse und Zielsetzung	Projektsponsor / Prozess-Experten (PE) / ML Experten (MLE)	Behörde des Projektsponsors
Phase 2: Proof of Concept	PE / MLE	Fach-Gremium / Test-Entwicklungs-Behörde(n) des UIS
Phase 3: Systemspezifikation	PE / MLE / Software-Entwickler	Alle Entwicklungs-Behörden der UIS
Phase 4: Umsetzung und Inbetriebnahme	PE / MLE / Software-Entwickler	Alle Entwicklungs-Behörden der UIS
Phase 5: Übergabe	PE, IT-Sicherheit	Rechenzentren der Behörden der UIS
Phase 6: Betrieb	Prozessbediener / PE / MLE / Software-Entwickler / IT-Sicherheit	Einsatz-, Entwicklungsbehörden und Rechenzentren der UIS

Abb. 2: Abbildung des Phasenmodells auf UIS

Abbildung 2 zeigt die Übertragung der Phasen auf UIS visuell. Die Hauptverantwortlichen einer Phase und die zuständigen Verwaltungseinheiten (z.B. die LAWA als Fachgremium mit Prozess-Entwicklern und ML Experten) sind den einzelnen Phasen zugeordnet.

### 5.3 Maschine Learning Pipeline Diagramm

Im folgenden wird beispielhaft ein Maschine Learning Pipeline Diagramm für ein Umweltinformationssystem definiert. Ein Ziel für die Nutzung eines ML-Algorithmus in einem UIS könnte die Vorhersage über die Entwicklung der Nitratbelastung an Messstellen sein. Das Ziel dieses Artikels ist die Diskussion einer möglichen Übertragbarkeit des ML4P Modells auf UIS, nicht die Entwicklung eines Vorhersageverfahrens, deswegen ist ein möglicher Algorithmus nur grob angerissen.

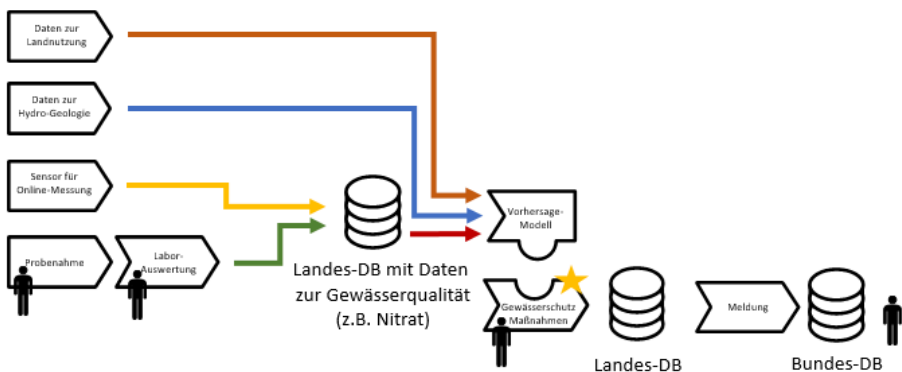


Abb. 3: Pipeline Diagramm

Abbildung 3 zeigt mögliche Eingabedaten:

- Geoinformationen aus unterschiedlichen Quellen, z.B. geologische Karten und Informationen über die Landnutzung (Ackerland, Grünland, etc.)
- Direkt gemessene Nitratwerte aus einzeln verfügbaren Sensoren.
- Produktiv gemessene Nitratwerte aus der Datenbank des Gewässerinformationssystems der letzten 20 Jahre, analysiert im Labor.

Die Eingabedaten speisen das Vorhersagemodell. Das Ergebnis des Algorithmus auf der rechten Seite der Abbildung sind aus dem Ergebnis der Vorhersage abgeleitete Gewässerschutzmaßnahmen. Die Prognoseergebnisse und Maßnahmen werden in die landeseigene Datenbank zurück gespielt und abschließend an die Datenbank der Bundesbehörde weitergeleitet.

Es ist in Planung das Vorgehensmodell an einem praktischen Projektbeispiel für die Erweiterung eines UIS mit Nitratprognosen zu testen. Im Rahmen dieses Projektbeispiels wird das Pipeline Diagramm während des Durchlaufs der Phasen des Vorgehensmodells kontinuierlich verfeinert und die für das Vorhersagemodell verwendeten konkreten KI-Methoden integriert.

#### **5.4 Virtuelle Prozessakte**

Die virtuelle Prozessakte des ML4P Vorgehensmodell beschreibt den Produktionskontext einer industriellen Anlage. Übertragen auf Umweltinformationssysteme bedeutet dies, dass in der virtuellen Prozessakte eines UIS alle relevanten Informationen bezüglich des ML-Algorithmus in der Prozessakte zusammengestellt werden. Dies sind insbesondere die Folgenden:

- Die grundlegende Bestandsdaten umfassen die Softwareversionen und Hersteller aller Komponenten des UIS und welche Daten und Funktionen die einzelnen Komponenten zur Verfügung stellen.
- Die erwartbaren Anlagendaten entsprechen dem zugrundeliegenden Datenmodell des UIS, sowie dessen Dokumentation.
- Weiterhin müssen Schnittstellen zwischen bestehenden Komponenten verschiedener Hersteller beschrieben werden. Die Schnittstellen sind auf viele unterschiedliche Arten realisiert und müssen entsprechend dokumentiert werden. Die Daten und zugehörigen Schnittstellen für den Zugriff auf die Daten können im Vorgehensmodell auf die Zusatzdaten abgebildet werden.

- Die zusätzlich erforderlichen Daten sind die, welche bisher nicht im UIS verfügbar sind, aber für die Lösung des zu implementierenden ML-Algorithmus erforderlich sind.

## 5.5 Eignung vorhandener ML4P Tools

Das ML4P Vorgehensmodell wird im Rahmen von Best-Practices durch die Entwicklung von technischen Komponenten unterstützt. Diese Tools können für die Realisierung einer ML Pipeline verwendet werden. Die bisher entwickelten ML4P-Komponenten basieren zumeist auf offener und frei verfügbarer Software. Diese Komponenten konzentrieren sich auf die Einbindung von Maschinen und Anlagen aus der Domäne der industriellen Produktion. Typische Standardschnittstellen für UIS unterscheiden sich allerdings zumeist von den in der Industrie eingesetzten Schnittstellen. Übliche UIS-Schnittstellen sind zum Beispiel:

- Direkter Austausch von Daten zwischen verschiedenen UIS Datenbanken einer Behörde via Datenbankaustauschtabelle.
- Im- und Export von Daten via Excel oder CSV-Dateien.
- Bedienung von proprietären Datenaustauschformaten (teilweise via XML-Schemata) von speziell für die Domäne Umwelt entwickelter Fach-Software. Typisches Beispiel ist der WasserBLiCK für die Berichterstattung gemäß der WRRL, der von den Gewässerinformationssystemen der Länder mit Daten beliefert wird. Weiterhin bestehen Schnittstellen zu allgemeiner proprietärer Software wie Laborinformationssystemen für die Integration von Analysedaten aus erhobenen Proben.
- Datenaustausch über offene Standard-Schnittstellen des Open Geospatial Consortiums (OGC). Ein Beispiel ist die Nutzung der OGC Sensor Things API. Dieser Standard ist sowohl für die Anbindung von Messdaten aus Umweltsensoren als auch als Schnittstelle für die Integration von zeitreihenbasierten Messdaten aus Umweltdatenbanken geeignet.

Für eine erfolgreiche technische Abbildung des ML4P Vorgehensmodells sollten für verbreitete Open Source Implementierungen, welche in den UIS eingesetzt werden, ML4P-Konnektoren entwickelt werden.

## 6 Ausblick

Da der Einsatz von Maschine Learning Methoden im industriellen Bereich weiter fortgeschritten ist als in der Domäne des Umweltschutzes wurde in diesem Artikel analysiert, inwieweit eine Methode aus der Industrie, nämlich das ML4P Vorgehensmodell für die

Einführung von ML in der Industrie, auf die Integration von KI-Methoden in bestehende behördliche Umweltinformationssysteme übertragbar ist.

Grundsätzlich ist eine an das ML4P angelehnte Vorgehensweise sehr sinnvoll, diese muss allerdings in bestimmten Bereichen für den Einsatz von behördlichen UIS angepasst oder erweitert werden. Dabei muss vor allem in der Phase 2 *Proof Of Concept* berücksichtigt werden, dass fachliche Lösungsansätze, die gesetzlichen Vorgaben entsprechen sollen, bisher häufig in behördenübergreifenden Gremien mit Fachexperten vor der Umsetzung in den im Einsatz empfindlichen UIS festgelegt werden. Für die erfolgreiche Entwicklung von ML-Algorithmen müssen bereits in dieser Phase ML-Experten herangezogen und Lösungsstrategien in geeigneten Test-UIS erarbeitet werden. Um eine angepasste Strategie für behördliche UIS zu entwickeln sind dabei die verschiedenen Phasen des industriellen Vorgehensmodells und die definierten Rollenkonzepte sehr hilfreich.

Als nächster Schritt ist der Test des Vorgehensmodells an einem praktischen Projektbeispiel in einem UIS sinnvoll. Im Rahmen eines Projektes können aus technischer Sicht bestehende ML4P-Konnektoren für den Einsatz in UIS getestet werden oder für UIS geeignete ML4P-Konnektoren entwickelt werden. Mit erfolgreicher Umsetzung kann so auf dem Machine Learning 4 Production Vorgehensmodell ein Machine Learning 4 Environment Modell (ML4E) aufsetzen.

## Literaturverzeichnis

- [DHvdS19] Desiree Hilbring, Robert Saenger, Jörg Stumpp; van der Schaaf, Hylke: Gewässerinformationssysteme auf Basis von XCNF und die Nutzung der Sensor Things API zur Veröffentlichung von Daten. AK-UIS, 2019.
- [Gr] Grambow, LAWA-Vorsitzender Herr Ministerialdirigent Prof. Dr. Martin: , Bund/Länder-Arbeitsgemeinschaft Wasser (LAWA). <https://www.lawa.de/>.
- [IS19] ISO: , ISO/IEC 10746-3:2009 Information technology — Open distributed processing — Reference model: Architecture — Part 3. <https://www.iso.org/standard/55724.html>, 2019.
- [JP20] Julius Pfrommer, Lars Wessels, Christian Frey Jürgen Beyerer: ML4P: Ein Standard-Vorgehensmodell für die Anwendung Maschinellen Lernens in der industriellen Produktion. In: Automatisierungskongress. 2020.
- [JS14] Jörg Stumpp, Desiree Hilbring, Thomas Gülden Anette Maetze: WaterFrame® - Neue Entwicklungen in den Gewässerinformationssystemen in Baden-Württemberg, Thüringen und Bayern. Fachdokumente Baden-Württemberg, 2014.
- [Mo17] Moßgraber, Jürgen: Ein Rahmenwerk für die Architektur von Frühwarnsystemen, Jgg. 29. KIT Scientific Publishing, 2017.
- [PE00] PARLAMENT, EUROPÄISCHES; EUROPÄISCHER, RAT: Richtlinie 2000/60/EG vom 23. Oktober 2000 zur Schaffung eines Ordnungsrahmens für Maßnahmen der Gemeinschaft im Bereich der Wasserpolitik–Wasserrahmenrichtlinie (WRRL). Amtsblatt der Europäischen Gemeinschaften, L, 327(1):22–12, 2000.

- [TJ19] Tobias Jetzke, Stephan Richter, Jan-Peter Ferdinand und Samer Schaat: , Künstliche Intelligenz im Umweltbereich. [https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-06-04\\_texte\\_56-2019\\_uba\\_ki\\_fin.pdf](https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-06-04_texte_56-2019_uba_ki_fin.pdf), 2019.
- [UB18] Usländer, Thomas; Batz, Thomas: Agile Service Engineering in the Industrial Internet of Things. MDPI Future Internet, 10(100):doi:10.3390/fi10100100, 2018.
- [Us05] Usländer, Thomas: Trends of environmental information systems in the context of the European Water Framework Directive. Environmental Modelling & Software, 20(12):1532–1542, 2005.
- [Us07] Usländer, Thomas: Reference Model for the ORCHESTRA Architecture (RM-OA). Version 2 (Rev 2.1). 2007.
- [WH00] Wirth, Rüdiger; Hipp, Jochen: CRISP-DM: Towards a standard process model for data mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. S. 29–39, 2000.