

TEXTE

79/2023

**Abschlussbericht**

# **Umsetzungskonzept umwelt.info – ein nutzer- und anwendungsorientiertes Angebot der Umweltverwaltungen**

**von:**

Dr. Matthias Bluhm, Georg Börner, Dr. Christopher Britsch, Ruth Busch, Dr. Udo Einspanier, Dr. Thore Fechner, Rico Illes, Sarah Lechler, Birgitta Lubahn, Rupert Pfeiffer, Lars Schneider, Manuel Schromm, Dr. Matthias Seuter, Dr. Uwe Voges

con terra GmbH, Münster

Dr. Johannes Franke,

Unabhängiges Institut für Umweltfragen e. V.

**Herausgeber:**

Umweltbundesamt



TEXTE 79/2023

Ressortforschungsplan des Bundesministeriums für  
Umwelt, Naturschutz und nukleare Sicherheit

Forschungskennzahl 3720 12 101 0  
FB001070

Abschlussbericht

## **Umsetzungskonzept umwelt.info – ein nutzer- und anwendungsorientiertes Angebot der Umweltverwaltungen**

von

Dr. Matthias Bluhm, Georg Börner, Dr. Christopher  
Britsch, Ruth Busch, Dr. Udo Einspanier, Dr. Thore  
Fechner, Rico Illes, Sarah Lechler, Birgitta Lubahn, Rupert  
Pfeiffer, Lars Schneider, Manuel Schromm, Dr. Matthias  
Seuter, Dr. Uwe Voges  
con terra GmbH, Münster

Dr. Johannes Franke,


Unabhängiges Institut für Umweltfragen e. V.

Im Auftrag des Umweltbundesamtes

## Impressum

### Herausgeber

Umweltbundesamt  
Wörlitzer Platz 1  
06844 Dessau-Roßlau  
Tel: +49 340-2103-0  
Fax: +49 340-2103-2285  
[buergerservice@uba.de](mailto:buergerservice@uba.de)  
Internet: [www.umweltbundesamt.de](http://www.umweltbundesamt.de)

 [/umweltbundesamt.de](https://www.facebook.com/umweltbundesamt.de)

 [/umweltbundesamt](https://twitter.com/umweltbundesamt)

### Durchführung der Studie:

con terra GmbH  
Martin-Luther-King-Weg 20  
48155 Münster

### Abschlussdatum:

November 2022

### Redaktion:

Fachgebiet I 1.7 Umweltinformationssysteme/-dienste, Satellitenfernerkundung,  
Dateninfrastruktur  
Marco Hohmann

Fachgebiet I 1.5 Nationale und internationale Umweltberichterstattung

Stefan Krämer

mit Unterstützung von

Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucher-  
schutz Referat T I 2 Umweltinformationen, Chief Data Officer, Künstliche Intelligenz  
Birte Solveig Kulla

Publikationen als pdf:

<http://www.umweltbundesamt.de/publikationen>

ISSN 1862-4804

Dessau-Roßlau, Juni 2023

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autorinnen und Autoren.

### **Kurzbeschreibung: Umsetzungskonzept umwelt.info – ein nutzer- und anwendungsorientiertes Angebot der Umweltverwaltungen**

Umweltbezogene Daten und Informationen sollen leicht auffindbar, gut zugänglich und möglichst frei verfügbar sein. Dafür führt das *Umweltbundesamt* ein Projekt zum *“Umwelt- und Naturschutz-Informationssystem für Deutschland” (umwelt.info)* durch. Ziel ist ein Web-Portal als zentraler Zugriffspunkt aller in Deutschland verfügbaren Daten und Informationen zu umweltbezogenen Themen. Die vorliegende Studie liefert ein Konzept zum Aufbau des Portals umwelt.info. Dieses Portal soll in seiner finalen Ausbaustufe die Metadaten zu Umweltdaten und -informationen aller Art umfassen. Hierbei sollen die Metadaten eine Prüfung auf Qualität, Aktualität sowie Vollständigkeit der beschriebenen Quellen erlauben und Informationen zur Rechtssicherheit von Daten liefern. Das Portal soll den Zugang zu Umweltdaten und -diensten gewährleisten und aggregierte und redaktionell aufbereitete Informationen zu ausgewählten Umweltthemen anbieten, womit eine Vielfalt an Zielgruppen angesprochen werden soll.

Das Umsetzungskonzept baut direkt auf der zuvor durchgeführten Machbarkeitsstudie [1] auf, die den inhaltlich-technologischen sowie den rechtlichen und organisatorischen Ist-Zustand bezüglich Umweltportalen in Deutschland aufzeigt. Die vorliegende Studie beinhaltet neben der technischen Konzeption für eine Basisumsetzung und ergänzende Quelladapter, die Erarbeitung eines agilen Vorgehensmodells sowie Vorschläge für die Organisation von Entwicklung und Betrieb von *umwelt.info*. Ergänzend werden vertragliche Aspekte im Hinblick auf die vorgesehene agile Entwicklung betrachtet und der aktuelle rechtliche Rahmen untersucht. Die Studie schafft die fachlichen, technischen und organisatorischen Grundlagen und liefert einen Vorschlag für die Realisierung des *umwelt.info*-Portals.

### **Abstract: Implementation concept of umwelt.info - a user and application-oriented offer for environmental management**

Data and information concerning the environment and nature protection should be easily findable, accessible and freely available. Thus, the *German Environment Agency* is realizing a project to implement an *environmental and nature conservation information system for Germany*. The goal is to create a web portal that grants central access to all in Germany existing data and information concerning the environment and nature protection. This study presents a concept for the implementation of this portal *umwelt.info*. The portal, once finished, will contain metadata of environmental data and information. The metadata should allow quality, relevance and integrity assessment of the integrated data sources and should give information about the legal certainty. The portal should grant access to environmental data and services and present aggregated and editorially processed information on selected environmental topics to accommodate a diverse audience.

This implementation concept is based on the previously conducted feasibility analysis, which has demonstrated the current state of environmental data portals concerning content, technology, legal framework and organization. The present study contains the technical concept for the basic implementation and additional source adapters, the development of an agile approach and prepositions for organizing implementation and operation of *umwelt.info*. Furthermore, the contractual aspects regarding agile development and the current legal framework are examined. This study creates the technical and organizational base and gives a preposition for implementing the *umwelt.info* portal.

## Inhaltsverzeichnis

Abbildungsverzeichnis.....	9
Tabellenverzeichnis .....	9
Abkürzungsverzeichnis .....	10
Zusammenfassung.....	12
Summary .....	16
1 Einleitung und Zielsetzung .....	20
2 Ziele und Bedarfe der Nutzenden .....	22
2.1 Ziele von umwelt.info .....	22
2.2 Nutzergruppen und User Stories .....	23
3 Systemarchitektur .....	27
3.1 Technische Randbedingungen und Qualitätsziele .....	27
3.2 Kontextabgrenzung.....	30
3.3 Lösungsstrategie .....	35
3.3.1 Harvesting und Crawling von Metadaten .....	35
3.3.2 Suche.....	36
3.3.3 Redaktionelle Inhalte einstellen .....	37
3.3.4 Metadaten manuell erfassen oder importieren.....	38
3.3.5 Daten nutzen / neue Daten ableiten .....	38
3.3.6 Nutzerverwaltung .....	38
3.3.7 Nutzerprofil.....	39
3.3.8 Programmierschnittstellen (API).....	40
3.3.9 Qualität von Metadaten .....	40
3.4 Komponenten .....	41
4 Anwendungsmöglichkeiten von Linked Data und KI-Methoden.....	43
4.1 Sucheingabevorschläge während der Eingabe .....	44
4.2 Ähnliche Suchen.....	44
4.3 Suche Erweitern .....	45
4.4 Ähnliche Datensätze .....	46
4.5 Antworten auf konkrete Fragen.....	47
4.6 Erweiterte Antworten / Mashups.....	47
4.7 Intelligente Suche .....	48
4.8 Qualitätssicherung von Metadaten .....	48
4.9 SPARQL-Client .....	49

4.10	Registrierung in der Linked Open Data Cloud .....	49
4.11	Weitere KI-Methoden .....	49
4.12	Workshopkonzept für begleitende Workshops während der Umsetzung .....	50
5	Agiler Prozess für umwelt.info .....	52
5.1	Agile Vorgehensmodelle .....	52
5.2	Anpassung an umwelt.info .....	54
5.3	Vertragsgestaltung .....	55
5.4	Planung der Stufen 1 und 2 von umwelt.info .....	56
6	Organisationskonzept.....	59
6.1	Beziehungen zu Anbietenden und Nutzenden .....	59
6.1.1	Beteiligung von Informationsanbietern.....	60
6.1.2	Beteiligung von Informationskonsumenten .....	61
6.2	Organisationseinheit umwelt.info .....	62
6.2.1	Entwicklung des Portals umwelt.info .....	62
6.2.2	Betrieb des Portals umwelt.info .....	63
6.2.3	Übergreifende Aufgaben .....	65
7	Aktualisierter Rechtsrahmen zur Umsetzung.....	66
7.1	Neuregelungen seit Erstellung der Machbarkeitsstudie.....	66
7.1.1	Umweltinformationsgesetz (UIG) .....	66
7.1.2	Urhebergesetz (UrhG).....	66
7.1.3	Telekommunikations-Telemedien-Datenschutz-Gesetz (TTDSG).....	66
7.1.4	Geologiedatengesetz (GeolDG) .....	66
7.1.5	Open Data: §12a EGovG .....	67
7.1.6	Datennutzungsgesetz (DNG).....	67
7.1.7	Gegenwärtige Regulierungsbestrebungen auf europäischer Ebene .....	67
7.2	Rechtliche Anforderungen an die Umsetzung von umwelt.info im Überblick .....	68
7.2.1	Keine rechtliche Verantwortlichkeit für verlinkte Inhalte .....	68
7.2.2	Anforderungen an die Beschaffung, Haltung und Verarbeitung von Metadaten .....	69
7.2.3	Verantwortlichkeit für redaktionelle Inhalte und selbst vorgehaltene Daten .....	71
7.2.4	Verantwortlichkeit für den allgemeinen Portalbetrieb .....	72
7.2.5	Verwaltungsvereinbarungen zur (Meta-)Datenbereitstellung .....	72
7.3	Handlungsempfehlungen.....	73
7.3.1	Verlinkung von Portalen und Datensammlungen Dritter .....	73
7.3.2	Redaktionelle Inhalte und Datenbereitstellung.....	74

7.3.3	Allgemeiner Portalbetrieb .....	75
7.3.4	Verwaltungsvereinbarungen zur (Meta-)Datenbereitstellung .....	75
8	Quellenverzeichnis .....	77
A	Anhang A: Systemarchitektur.....	81
B	Anhang B: Vorgehensmodell .....	82
C	Anhang C: Glossar.....	83



## Abbildungsverzeichnis

Abbildung 1:	User Journey Map: Beratende Ingenieur*innen .....	24
Abbildung 2:	User Journey Map: Politikberater*in auf lokaler Ebene .....	25
Abbildung 3:	User Journey Map: Mitarbeiter*in Datenmanagement öffentliche Verwaltung.....	25
Abbildung 4:	User Journey Map: Lehrer*innen .....	25
Abbildung 5:	Kontextabgrenzung des umwelt.info-Portals.....	31
Abbildung 6:	Kontextabgrenzung (technisch) des umwelt.info- Umweltportals .....	33
Abbildung 7:	Genereller Ablauf beim Harvesting und Crawling .....	36
Abbildung 8:	Verschiedene Ebenen in der Projektstruktur .....	55
Abbildung 9:	Entwurf einer Roadmap .....	56

## Tabellenverzeichnis

Tabelle 1:	Technische Randbedingungen der IT des Bundes .....	27
Tabelle 2:	Qualitätsziele .....	29
Tabelle 3:	Kommunikationsbeziehungen des umwelt.info Portals (Fachlicher Kontext).....	31
Tabelle 4:	Input und Output des umwelt.info Portals (Technischer Kontext) .....	34

## Abkürzungsverzeichnis

Abkürzung	Bedeutung
AG	Auftraggeber*in
AN	Auftragnehmer*in
API	Application Programming Interface (Programmierschnittstelle)
BMUV	Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz
CMS	Content Management System
CPU	Central Processing Unit
CRUD	Create/Read/Update/Delete
CSW	Catalogue Service Web
CUD	Create/Update/Delete
CS	Catalogue Service
DB	Datenbank
DCAT	Data Catalogue Vocabulary
DCAT-AP	Data Catalogue Vocabulary-Application Profile for data portals in Europe
DCAT-AP.DE	Data Catalogue Vocabulary-Application Profile for data portals in Europe, deutsches Profil
DevOps	Zusammensetzung aus Development (Dev) und Operations (Ops)
DGA	Data Governance Act
DNG	Datennutzungsgesetz
DS	Daten Service
DSGVO	Datenschutzgrundverordnung
DU	Unstrukturierte Daten
EGovG	E-Government Gesetz, Open Data Gesetz
GeolDG	Geologiedatengesetz
GSB	Government Site Builder
HTML	Hypertext Markup Language
INSPIRE	Infrastructure for Spatial Information in the European Community
ID	Identification
ISO	Internationale Organisation für Normung
JSON	JavaScript Object Notation
KI	Künstliche Intelligenz
LD	Linked Data
LDAP	Lightweight Directory Access Protocol
MD	Metadaten
MQA	Metadata Quality Assessment

Abkürzung	Bedeutung
OGC	Open Geospatial Consortium
PHP	Hypertext Preprocessor
RAM	Random-Access Memory
RDF	Resource Description Framework
REST	Representational State Transfer
RS	Recommender System
SDW	Spatial Data on the Web
SSO	Single Sign-On
TTDSG	Telekommunikations-Telemedien-Datenschutz-Gesetz
UIG	Umweltinformationsgesetz
UBA	Umweltbundesamt
UI	User Interface
UrhG	Urhebergesetz
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
UX	User Experience
WMS	Web Map Service
XP	Extreme Programming
XML	Extensible Markup Language

## Zusammenfassung

Umweltbezogene Daten und Informationen bilden das Fundament nachhaltiger Entwicklung und müssen leicht auffindbar, gut zugänglich und möglichst frei verfügbar sein. Vor diesem Hintergrund soll das **Web-Portal *umwelt.info*** mit dem Ziel aufgebaut werden, als zentraler Zugriffspunkt alle bundesweit öffentlich verfügbaren Daten und Informationen zu umweltbezogenen Themen an einem Ort zusammenzuführen. Die Daten und Informationen sollen zielgruppengerecht aufbereitet allen Bürger\*innen, Wissenschaft, Forschung und Bildung, Unternehmen und Industrie sowie der Verwaltung zur Verfügung stehen. Das Portal *umwelt.info* soll es ermöglichen, auch auf Informationen zuzugreifen, die zuvor nur schwierig zu finden und schwierig zugänglich waren.

Das *umwelt.info*-Portal soll umfangreiche Suchfunktionen und den Zugang zu Metadaten über Daten, Dienste, Berichte, Forschungsergebnisse, Gutachten, Bildungsmaterialien ebenso wie zu Rechts- und Verwaltungsvorschriften, Förderprogrammen oder Verfahren der Umwelt- und Naturschutzverwaltungen aller administrativen Ebenen bieten. Damit soll für mehr **Transparenz** der Verwaltung gegenüber der allgemeinen Öffentlichkeit und für eine verbesserte, innovative Nutzung von Umweltdaten und -informationen gesorgt werden. Das Portal soll auch zur besseren **Erfüllung von EU-Richtlinien** beitragen, insbesondere der Umweltinformationsrichtlinie (2003/4/EG), welche die Vorgaben der Aarhus-Konvention für das Recht auf freien Zugang zu Umweltinformationen auf EU-Ebene umsetzt. In *umwelt.info* werden Daten und Informationen auffindbar und zugänglich gemacht, jedoch keine Primärdaten gehalten, sodass Fachinformationssysteme, die auf allen Verwaltungsebenen existieren, nicht ersetzt werden. Für die Umsetzung wurde als übergeordnete Zielsetzung folgende Vision entwickelt:

### Vision für *umwelt.info*:

- ▶ Ein Portal, über das alle deutschen Umwelt- und Naturschutzinformationen erreicht werden können;
- ▶ Ein Portal, das zu allen Umwelt- und Naturschutzinformationen in Deutschland Metadaten liefert, welche auf Qualität, Aktualität sowie Vollständigkeit prüfbar sind und Informationen zur Rechtssicherheit von Daten liefern;
- ▶ Ein Portal, das den offenen Zugang zu Umweltdaten und -diensten gewährleistet;
- ▶ Ein Portal, in dem Informationen mit Umwelt- und Naturschutzbezug redaktionell aufbereitet werden;
- ▶ Ein Portal für unterschiedliche Zielgruppen von Fachleuten bis hin zu interessierten Privatpersonen;
- ▶ Ein Portal, das Kommunikation und Zusammenarbeit fördert.

Das vorliegende Umsetzungskonzept setzt direkt auf der zuvor durchgeführten **Machbarkeitsstudie** [1] auf. Diese zeigt den inhaltlich-technologischen sowie den rechtlichen und organisatorischen Ist-Zustand bezüglich Umweltportalen in Deutschland auf. Außerdem beinhaltet sie eine umfangreiche Analyse der Zielgruppen und deren Bedarfe. Auf Basis der Nutzeranforderungen wurde in der Machbarkeitsstudie der grobe Entwurf einer Systemarchitektur für *umwelt.info* erstellt und eine Umsetzung in vier Ausbaustufen vorgeschlagen: Basisumsetzung (1), Ergänzende Quelladapter (2), Erweiterung um Linked Data

(3) und Künstliche Intelligenz (4). Sowohl die Machbarkeitsstudie als auch das vorliegende Umsetzungskonzept wurden durch einen **Projektbeirat** begleitet, dem Vertreter\*innen von Landes- und Bundesbehörden bzw. Ministerien angehörten. Mit dem Projektbeirat wurden mehrere Besprechungen und Informationsveranstaltungen durchgeführt, sodass die Mitglieder ihre Meinungen und Erfahrungen in die Konzepterstellung einbringen konnten. Insbesondere wurde mit dem Projektbeirat auch die zukünftige Einbindung datenhaltender Stellen und der Nutzenden besprochen. Einzelne Beiratsmitglieder nahmen auch an Projektbesprechungen und Workshops teil, z. B. zum agilen Vorgehen oder zum Thema Künstliche Intelligenz.

Ein wesentliches Ergebnis des Umsetzungskonzeptes bildet die detaillierte Beschreibung der **Systemarchitektur**. Diese wurde methodisch und inhaltlich in Anlehnung an das Arc42 Architektur-Template [7] erstellt, das eine Vorlage für die Entwicklung, Dokumentation und Kommunikation von Software- und Systemarchitekturen bietet. Wesentliche Zielgruppen der Beschreibung der Systemarchitektur sind Personen, die an der Umsetzung von *umwelt.info* arbeiten oder die an Entscheidungen über das System und dessen Entwicklung beteiligt sind. Dies sind insbesondere das Entwicklungsteam, die Projektleitung und das Betriebsteam. Neben den Anforderungen der Nutzenden bilden die technischen Randbedingungen sowie die Qualitätsziele die Basis für die Systemarchitektur. Die **technischen Randbedingungen** orientieren sich an der Architekturrichtlinie der IT des Bundes [5]. Zu den **Qualitätszielen** zählen insbesondere eine hohe Qualität der Metadaten, die einfache Aktualisierbarkeit und Testbarkeit, eine gute Skalierbarkeit sowie eine hohe Sicherheit, Verfügbarkeit und Zuverlässigkeit. Das Architekturkonzept beschreibt ein System aus den folgenden **Hauptkomponenten**: Content Management System (CMS), Metadaten-Harvesting und -Crawling, Data-Check-In, Metadaten-Index, Application, Programmierschnittstellen, Identity Management, Metadata Quality Assessment, Recommender System und Nutzertracking. Die genannten zehn Hauptkomponenten werden in ca. **50 Subkomponenten** untergliedert und beschrieben. Die Entwicklung der Architektur orientierte sich insbesondere an folgendem **Entwurfsmuster**: Eine lose Kopplung von Komponenten bei einem gleichzeitigen starken Zusammenhalt führen zu einer gut wartbaren und testbaren Architektur. Dieses Konzept kann durch Microservices realisiert werden. **Microservices** stellen unabhängige Funktionen des Systems dar, die jeweils eine bestimmte Teilaufgabe erledigen und miteinander über sprachunabhängige Programmierschnittstellen kommunizieren.

Über ein **Content Management System (CMS)** werden der Einstiegspunkt des Portals für Anwendende und die Suche nach Daten realisiert. Es stellt auch die Nutzeroberfläche für Redakteur\*innen dar, um redaktionelle Inhalte zu erstellen. Die Komponente **Metadaten-Harvesting und -Crawling** ermittelt aus den verschiedenen Datenquellen die Metadaten, indiziert diese und stellt sie im Portal *umwelt.info* zur Verfügung. Dieser Prozess ist in drei Schritte unterteilt. Im ersten Schritt werden die Metadaten aus den verschiedenen Quellen extrahiert, anschließend erfolgt die Transformation der Metadaten in das Zielformat und abschließend werden die Daten in den Metadaten-Index geschrieben. Zusätzlich wird es eine Komponente zur Steuerung dieses Prozesses geben. Damit wird geregelt, wann für eine bestimmte Datenquelle der Prozess gestartet werden soll und welche Harvester bzw. Crawler zu verwenden sind. Der **Metadaten-Index** enthält alle Metadaten, die über das Harvesting und Crawling integriert oder manuell eingepflegt wurden. Der Metadaten-Index stellt einen effizienten Indizierungs- und Suchmechanismus bereit und ist in der Lage, das verwendete Metadatenmodell zu interpretieren und eine Indizierung über ausgewählte Suchparameter oder als Volltext vorzunehmen. Diese Funktionen können von einer der verbreiteten Suchmaschinen übernommen werden, indem diese in *umwelt.info* integriert wird. Im *umwelt.info*-Portal sind je nach zu indizierender Datenquelle für das Harvesting bzw. Crawling der Metadaten

unterschiedliche Zugriffsmechanismen nötig. Harvester und Crawler sind beispielsweise für standardisierte Metadaten- oder Datendienste (z. B. Dienste des Open Geospatial Consortiums, und INSPIRE-Dienste), Dateiverzeichnisse, Datenbanken und Web-Anwendungen vorgesehen. Für die Erhebung von Statistiken und Prüfung der Qualität von indizierten Metadatensätzen wird eine **Metadata Quality Assessment**-Komponente eingesetzt.

Die Nutzung des **Linked Data**-Ansatzes und von Methoden und Konzepten aus dem Bereich der **Künstlichen Intelligenz** (KI) wurden geprüft, um die Nutzbarkeit des Portals in kommenden Ausbaustufen zu erweitern. Insbesondere sollen die Suchergebnisse zu den Sucheingaben der Nutzenden optimiert und intelligente Hilfestellungen gegeben werden. Dies kann dadurch erfolgen, dass während der Suche passende Vorschläge angezeigt werden, dass die Suche durch Anzeige alternativer Begriffe erweitert wird, dass ähnliche Datensätze als Empfehlung in der Ergebnisliste einer Suche angezeigt werden oder dass die passenden Informationen auf konkrete Fragen mittels einer „Question-Answering“-Komponente geliefert werden. Als Ansatzpunkte für KI-Methoden in der Metadaten-Qualitätssicherung wurden zwei Bereiche herausgearbeitet: Unterstützungen während des Ausfüllens von Metadaten-Formularen und Überprüfung bestehender Metadaten. Über die Prüfungen auf technische oder formale Korrektheit hinaus sollen weitere inhaltliche Fehler erkannt werden. Dabei können Methoden aus dem Bereich der Natürlichen Spracherkennung eingesetzt werden, um Inhalte unterschiedlicher Felder untereinander und Metadaten mit den Daten selbst abzugleichen.

Im Zuge der Erarbeitung der Systemarchitektur sowie deren Erweiterung um den Themenkomplex Künstliche Intelligenz wurden die Anwenderanforderungen und die **Designentwürfe** aus der Machbarkeitsstudie überarbeitet. Unter Anwendung eines User-Centered-Design Ansatzes wurden die bisher sechs **Personas** um die Persona „Redakteur\*in“ erweitert und es wurden vier User Journeys beschrieben. Damit werden exemplarisch **Nutzungsszenarien** der Zielgruppen „Beratende Ingenieur\*in“, „Politikberater\*in auf lokaler Ebene“, „Mitarbeiter\*in im Datenmanagement öffentliche Verwaltung“ und „Lehrer\*in“ dargestellt.

Im Umsetzungskonzept wurden auch die zukünftigen **Organisations- und Projektstrukturen** erarbeitet. Als erster Umsetzungsschritt wurde bereits das Nationale Zentrum für Umwelt- und Naturschutzinformationen als Außenstelle des Umweltbundesamtes in Merseburg eingerichtet. Die **Aufgaben des Zentrums** während der Realisierung des Portals sind die Steuerung des gesamten Entwicklungsprozesses, die Spezifikation von Anforderungen und die Kommunikation mit datenhaltenden Stellen. Während des Betriebs des Portals sind der technische Betrieb zu organisieren, Metadatensätze einzustellen und deren Qualität zu gewährleisten, die Funktionalitäten zu prüfen, Nutzende zu verwalten, Fachinhalte zu erstellen und im Portal verfügbar zu machen, Informations anbietende zu beraten sowie Gremien- und Öffentlichkeitsarbeit durchzuführen. Da die weitere Planung einen schrittweisen Aufbau des Portals vorsieht, werden sich die Aufgaben aus der Realisierungs- und Betriebsphase überlappen. Neben den Aufgabenbeschreibungen für die Organisationseinheit *umwelt.info* werden auch alle weiteren Beteiligten bzw. Betroffenen berücksichtigt und deren Rollen und Aufgaben beschrieben. Die **Beteiligung der daten- und informationshaltenden Stellen**, aber auch anderer Stakeholder soll über unterschiedliche Wege erfolgen, wie z. B. Gremien, Informationsveranstaltungen und direkte Beratungsangebote durch *umwelt.info*. Für die **Beteiligung der Nutzenden** werden Optionen wie Informationsveranstaltungen, Workshops der Nutzenden, ein Beteiligungsportal oder Online-Konsultationen genannt.

Entwicklung und Betrieb von *umwelt.info* können als ein komplexes Projekt angesehen werden. Die Komplexität ergibt sich insbesondere aus einer heterogenen Nutzergruppe, sodass für das Portal *umwelt.info* ein Zielraum definiert werden kann, aber zu erwarten ist, dass sich die Anforderungen im Projektverlauf noch ändern werden. Eingebettet ist *umwelt.info* zudem in ein

komplexes organisatorisches und politisches Umfeld, das alle Verwaltungsebenen von der Kommune bis hin zur Europäischen Union umfasst. Zudem sind die rechtlichen Rahmenbedingungen sehr divers. Dies lässt ein **agiles Vorgehensmodell** besonders geeignet für die Entwicklung des Vorhabens erscheinen. Drei agile Vorgehensmodelle wurden identifiziert, die als Basis für eine Softwareentwicklung geeignet erscheinen: Scrum, Kanban und Extreme Programming. Diese Vorgehensmodelle besitzen einen hohen Bekanntheitsgrad und haben sich in der Praxis bewährt, sodass vielfach Kenntnisse bereits vorhanden bzw. Informationen leicht zu beschaffen sind. Eine vergleichende Bewertung dieser Modelle erfolgte unter anderem anhand der Kriterien Abdeckung des Prozesses, Anpassbarkeit, Stabilität gegenüber personellen Veränderungen und Einarbeitungsaufwand. Als Grundlage für die Realisierung von *umwelt.info* wurde **Scrum** als am besten geeignet bewertet. Da das Scrum-Modell jedoch nicht alle Anforderungen für *umwelt.info* vollständig abdeckt, wurden die Regeln und Inhalte dieses Modells entsprechend den Anforderungen von *umwelt.info* angepasst und detailliert beschrieben. Zwecks weiterer Planung wurden Entwürfe der Roadmap und des Product-Backlogs erarbeitet. Die Roadmap weist eine Reihenfolge von Arbeitspaketen aus, auf deren Grundlage die Definition der Sprints erfolgen kann. Sie basiert auf den Komponenten der Systemarchitektur und stellt einen Weg für die Basisumsetzung und ergänzende Quelladapter dar. Ein agiles Vorgehen bei der Projektumsetzung erfordert die Schaffung geeigneter vertraglicher Grundlagen. Aus diesem Grund wurde ein **Muster für eine Rahmenvereinbarung** entworfen, die auf Standardverträgen des Umweltbundesamtes basiert und an ein agiles Vorgehensmodell angepasst wurde. Das Muster kann dem Umweltbundesamt als Basis dienen, um konkrete Verträge mit Dienstleistern zu schließen.

Die **rechtlichen Rahmenbedingungen** für den Aufbau eines Umwelt- und Naturschutzinformationssystems wurden in der Machbarkeitsstudie [1] bereits ausführlich betrachtet. Im vorliegenden Umsetzungskonzept werden Änderungen im Rechtsrahmen nach Abschluss der Machbarkeitsstudie kurz dargestellt und eingeordnet. Dies betrifft beispielsweise das Umweltinformationsgesetz, das Urhebergesetz, das neue Geologiedatengesetz sowie Entwürfe für einen Data Act und einen Data Governance Act auf EU-Ebene. Es wird begründet, dass die bloße Verlinkung von Inhalten weder eine urheberrechtlich relevante Handlung noch eine Verarbeitung personenbezogener Daten bedeutet. Die derzeitige Konzeption sieht vor, dass das Portal als Metadatensuchmaschine angelegt wird, also über Hyperlinks auf externe Inhalte verweist. Die Haltung von Inhalten und Daten liegt weiterhin in der Verantwortung der datenhaltenden Stellen, sodass *umwelt.info* als portalbetreibende Stelle hierfür grundsätzlich keine rechtliche Verantwortung hat. Eine rechtliche Verantwortlichkeit besteht jedoch, wenn auf *umwelt.info* eigene redaktionelle Inhalte bereitgestellt werden. Zu beachten ist bei der Gestaltung dieser redaktionellen Inhalte vor allem das Urheberrecht. Abschließend werden rechtliche Empfehlungen bzw. Leitlinien für das Betriebsteam von *umwelt.info* formuliert. Konkret betrifft dies die Verlinkung von externen Datenbeständen, die Bereitstellung von Daten und Inhalten unmittelbar auf *umwelt.info*, den allgemeinen Portalbetrieb sowie den Abschluss von Verwaltungsvereinbarungen zum Datenaustausch.

Im Umsetzungskonzept werden in sieben Kapiteln die Ziele von *umwelt.info* (1), die Nutzerbedarfe (2), die Systemarchitektur (3), die Anwendungsmöglichkeiten künstlicher Intelligenz (4), der agile Prozess für Entwicklung und Betrieb (5), die Organisation (6) und rechtliche Aspekte (7) beschrieben. Weiterhin werden detaillierte Beschreibungen zur Systemarchitektur (Anhang A) und zum agilen Vorgehensmodell (Anhang B) bereitgestellt.

Mit dem vorliegenden Konzept wurden die fachlichen, technischen und organisatorischen Grundlagen geschaffen und es liegt ein Vorschlag für die Realisierung des *umwelt.info*-Portals vor.



## Summary

Environmental data and information are the base for sustainable development and must be easily findable, accessible and freely available. Therefore, the *German Environment Agency* implements the **online portal *umwelt.info*** aiming to become the central access point to all nationwide open data and information on environmental topics. The data and information will be processed to fit the needs of target groups: citizens, scientists, research and education, business and industry, administration and government agencies. The portal will enable access to information that was hard to find or access before.

The portal *umwelt.info* will provide extensible search options and access to metadata of data, services, reports, research results, expert reports, educational materials as well as legal and administrative regulations, funding programs, or procedures of environmental and nature conservation administrations at all administrative levels. This portal should provide more **transparency** between the administration and the general public and should enable an enhanced, innovative use of environmental data and information. Furthermore, *umwelt.info* should contribute to a better **fulfilment of European guidelines**, for example, the directive on public access to environmental information (2003/4/EG), which implements the requirements of the Aarhus Convention for the right to free access to environmental information at the EU level. Data and information will be made accessible on *umwelt.info*. However, *umwelt.info* will not store any primary data and will not substitute existing information systems on the administrative level. The following vision is the primary goal for the implementation.

### Vision for *umwelt.info*:

- ▶ A portal which provides access to all German environmental and nature conservation data and information.
- ▶ A portal, which provides metadata to all environmental and nature conservation data and information in Germany. These metadata allow the assessment of quality, relevance and integrity and give information about the legal certainty.
- ▶ A portal that grants open access to environmental data and services.
- ▶ A portal which provides editorial prepared information about environmental and nature conservational topics
- ▶ A portal for diverse target groups from professionals to interested private individuals
- ▶ A portal enhancing communication and cooperation

This concept is based on a previously conducted **feasibility study** [1] which demonstrates the current state of environmental online portals in Germany concerning content, technological implementation, juridical framework and organization. It also includes a comprehensive analysis of the target groups and their needs. Based on the user requirements, a rough draft of a system architecture for *umwelt.info* was created and an implementation in four expansion stages was proposed: basic implementation (1), supplementary source adapter (2), extension with linked data (3) and artificial intelligence (4). Both the feasibility study and the present implementation concept were accompanied by a **project advisory board**, which included representatives of state and federal authorities and ministries. Informative meetings and workshops were held with the project advisory board and the members could contribute to the concept draft. In particular, the future engagement of data providers and users was discussed with the project



advisory board. Some advisory board members also took part in project meetings and workshops, e.g., on agile procedures or about artificial intelligence.

A significant result of the implementation concept is the detailed description of the **system architecture**. This document was created based on the Arc24 architecture template [7], which is a template for the development, documentation, and communication of software and system architecture. The system architecture is mainly written for persons that work on the implementation of *umwelt.info* or make decisions about the system and its implementation. Specifically, these are the development team, the project leader, and the operating team. Besides the user requirements, the **technical constraints**, according to the architectural guidelines of the IT of the federal government, and the **quality goals** are the base for the system architecture. The quality goals include high-quality metadata, simplified maintenance and easy testability, good scalability as well as high levels of security, availability and reliability. The concept of the system architecture is described by the following **main components**: Content Management System, Metadata-Harvesting and Crawling, Data-Check-In, Metadata Index, Application, Implementing APIs, Identity Management, Metadata Quality Assessment, Recommender System und User Tracking. These main components are divided and described within 50 subcomponents. This concept can be realized by using **Microservices** – independent functions of the system completing different subtasks and communicating with each other language independent via APIs.

The **Content Management System (CMS)** is the user's access point to the portal and the data search. Additionally, it is the user interface for editors to create editorial content. The **Metadata-Harvesting and Crawling** component gathers metadata from different data sources. This process is separated into three steps. Firstly, the metadata are extracted from different sources, secondly, the metadata are transformed to fit the target format and lastly, the metadata are written into the Metadata Index. Moreover, there will be a component to manage this process that helps control the start of the process for each individual data source and which Harvester or Crawler to use. The **Metadata Index** contains all metadata integrated by the Harvester or Crawler or added to the system manually. This component is an efficient indexing and search mechanism which can interpret the metadata model and create indexing for selected search parameters or full-text search. One of the well-known search engines can adopt this function when integrated in *umwelt.info*. In the portal, different accessing mechanisms are necessary depending on the index data source of the Harvesting or Crawling. Harvester and Crawler, for example, are used for standardized metadata and data services (e. g. Services from Open Geospatial Consortiums, or INSPIRE-Services) file directories, databases, and web applications. To create statistics and review the indexed metadata, a **Metadata Quality Assessment** component is inserted.

The usage of **Linked Data** approaches and **Artificial Intelligence (AI)** methods and concepts were examined to expand the usability of the portal in later expansion stages. In particular, optimized search results and an intelligent assistance should be given by displaying suitable suggestions during the search input, expanding the search, displaying alternative terms, recommending similar data records in the search results, or respond to specific questions using a Question-Answering component. Relevant optimization tasks were identified for metadata quality assessment through AI methods for filling out metadata forms and checking existing metadata. In addition to the checks for technical or formal correctness, contextual errors in content should be also detected. Methods from the field of natural language recognition can be used to compare the entries of different metadata items.

The users' requirements and the **user interface sketches** from the feasibility study were

revised with the help of the developed system architecture and ideas for AI extension. For example, the “editor” expands the group of personas in the user-centered design approach. Furthermore, four user journeys were generated describing **application scenarios** of the personas “consulting engineer”, “political consultant at local level”, “employee in data management in public administration” and “teacher”.

The future **organizational project structures** were developed in the implementation concept. Therefore, the *German Environment Agency* build the new branch office in Merseburg called the *National Center for Environmental and Nature Conservation Information*. Its implementation tasks are the control of the entire development process, the specification of requirements and the communication with data-holding agencies. The operating tasks are the organization of the technical operation, metadata integration, metadata quality assessment, function tests in the operating system, user management, editorial content creation, editorial content management, consultation of data- and information-holding providers, governance management and public relations. Since further planning envisages a step-by-step construction of the portal, the tasks from the implementation and operation phases will overlap. In addition to the task descriptions for the new *National Center*, all other participants or those affected are also considered and their roles and tasks are described. The data- and information-holding agencies and important stakeholders should participate in *umwelt.info* through, e.g., committees, conferences, and direct advisory services. Whereas, users should participate through, e.g., information events, user workshops, activity in a participation portal or online consultations.

The complexity of implementing and operating *umwelt.info* is affected in particular by heterogeneous target groups, which also define functions of the portal. However, adaptations can be expected as the requirements for target groups will change during the project. Furthermore, *umwelt.info* is also embedded in a complex organizational and political environment that encompasses all administrative levels from local authorities to the European Union. In addition, the legal framework conditions are very diverse. An **agile process model** seems particularly suitable for the implementation of the project. Three suitable agile process models were comparatively evaluated as a basis for software development: Scrum, Kanban, and Extreme Programming. These process models are well known and well established. Thus, information on the models were easy to obtain. The evaluation of models was based on criteria such as the process, adaptability, stability in the face of personnel changes and training effort. Scrum was identified as the most suitable basis for the realization of *umwelt.info*. However, the concept describes suggestions for adaptations of the **Scrum** model to cover all requirements for *umwelt.info*. Drafts of the roadmap and product backlog were developed for planning of the first implementation stages. The roadmap is based on the components of the system architecture and provides a path for the basic implementation and supplementary source adapters. An appropriate contractual foundation is necessary for the agile approach of the project implementation. Therefore, a sample **framework agreement** was drafted based on standard contracts of the *German Environment Agency* and adapted to an agile procedure model. The sample can serve as a basis to conclude contracts with service providers.

The feasibility study [1] has already considered the **legal framework** for setting up an information system concerning the environment and nature conservation. The present implementation concept presents changes in the legal framework after completion of the feasibility study briefly. This concerns for example the Environmental Information Act, the Copyright Act, the new Geological Data Act, and drafts for a Data Act and a Data Governance Act at the EU level. A main implication is that the mere linking of content does not constitute an act relevant under copyright law, nor does it constitute the processing of personal data. The portal will contain a metadata search engine, which, i.e., will refer to external content via hyperlinks.

The maintenance of content and data remains the responsibility of the data providers, whereas *umwelt.info* is legally not responsible as the portal-operating body, except *umwelt.info* provides its own editorial content. Above all, copyright law must be observed in the design of this editorial content. Legal recommendations or guidelines are formulated for the *umwelt.info* operating team as an outcome of this concept. Specifically, this concerns the linking of external databases, the provision of data and content directly on *umwelt.info*, general portal operation, and the conclusion of administrative agreements for data exchange.

In the implementation concept, seven chapters describe the goals of *umwelt.info* (1), the user needs (2), the system architecture (3), the application possibilities of artificial intelligence (4), the agile process for development and operation (5), the organization (6), and legal aspects (7). Furthermore, detailed descriptions of the system architecture (Appendix A) and the agile process model (Appendix B) are provided.

With the following concept, the functional, technical and organizational foundations have been created and a proposal for the realization of the *umwelt.info* portal is presented.

# 1 Einleitung und Zielsetzung

Umweltbezogene Daten und Informationen sollen leicht auffindbar, gut zugänglich und möglichst frei verfügbar sein. Das Umweltbundesamt entwickelt dafür ein „Umwelt- und Naturschutz-Informationssystem für Deutschland“ (umwelt.info)<sup>1</sup>. Ziel ist es, einen zentralen Zugriffspunkt aller bundesweit öffentlich verfügbaren Daten und Informationen zu umweltbezogenen Themen aufzubauen - das Web-Portal umwelt.info.

## Definition Daten und Informationen mit Umwelt- und Naturschutzbezug

Im Umsetzungskonzept werden oft „Umweltdaten“ oder „Daten und Informationen mit Umweltbezug“ genannt. Hierbei sind „Naturschutzdaten“ oder „Daten und Informationen mit Naturschutzbezug“ inbegriffen. Aus Gründen des Leseverständnisses werden unter dem Oberbegriff „Umweltdaten“ sowohl Daten als auch Informationen verstanden. „Umweltdaten“ werden in umwelt.info durch Metadaten abgebildet.

Das vorliegende Umsetzungskonzept setzt direkt auf der zuvor durchgeführten Machbarkeitsstudie [1] auf. Dort wurde herausgearbeitet, in welcher Form langfristig ein neues Fach- und Ebenen-übergreifendes Angebot (Bund, Länder, Kommunen, Wissenschaft, Wirtschaft etc.) mit Zugang zu Daten, Diensten, Berichten, Forschungsergebnissen, Gutachten, Bildungsmaterialien ebenso wie zu Rechts- und Verwaltungsvorschriften, Förderprogrammen oder Verwaltungsverfahren der Umwelt- und Naturschutzverwaltungen machbar und nützlich sein könnte. Die Inhalte von umwelt.info sollen zudem dergestalt aufbereitet und zugänglich sein, dass sie auch den Einsatz von KI-Verfahren vereinfachen und befördern. Zugleich wird die Teilhabe Dritter an einer Verbesserung der Datengrundlagen ermöglicht und es werden neuen Akteuren Wege eröffnet, ihre Beiträge zum Klima-, Umwelt- und Naturschutz qualitätsgesichert sichtbar zu machen. Auch Anlass und Hintergründe, die dazu führten, dieses Projekt weiter voranzutreiben, wurden in der Machbarkeitsstudie dargestellt.

Die Machbarkeitsstudie [1] beinhaltet die folgenden wesentlichen Aspekte:

- ▶ Eine Auflistung und Kurzbeschreibung bestehender Portale aus dem Umweltbereich sowie ausgewählter übergreifender Portale. Darauf aufbauend wird der inhaltlich-technologische Ist- und Planungsstand aufgezeigt. Dies umfasst eine Übersicht bekannter Umwelt- und Informationsportale sowie eine Darstellung der bereitgestellten Daten ausgewählter Portale und die Vorstellung eines Best-Practice Beispiels anhand von Wasser-DE (siehe Kap. 2.2 in [1]);
- ▶ Eine Erörterung des rechtlichen Ist-Stands, u.a. hinsichtlich der Übertragbarkeit und Bereitstellungsverpflichtungen, Anforderungen seitens des Behindertengleichstellungsgesetzes und Wahrung des Urheberrechts (siehe Kap. 2.3 in [1]);
- ▶ Eine Darstellung des organisatorischen Ist-Stands (siehe Kap. 2.4 in [1]);
- ▶ Eine umfangreiche Analyse der Zielgruppen und Bedarfe der Nutzenden (siehe Kap. 3 in [1]);

<sup>1</sup> In diesem Dokument wird für das beschriebene System die aktuelle Bezeichnung ‚umwelt.info‘ anstelle des ehemaligen Arbeitstitels ‚UNIS-D‘ verwendet. ‚UNIS-D‘ wird jedoch noch in vielen referenzierten Dokumenten, z. B. der Machbarkeitsstudie [1], genutzt. Die betreibende Stelle, bisher mit ‚UNIZ-D‘ bezeichnet, wird jetzt ‚Nationales Zentrum für Umwelt- und Naturschutzinformationen‘ (umwelt.info) benannt.

- ▶ Soll-Vorschläge, die aufbauend auf den Bedarfen der Nutzenden erstellt wurden und deren Machbarkeit im Rahmen von Workshops evaluiert wurde (siehe Kap. 4.2./4.3 in [1]);
- ▶ Ein Vorschlag für eine Umsetzung in vier Ausbaustufen (siehe Kap. 4.4 in [1]):
  - Ausbaustufe 1 – Basisumsetzung
  - Ausbaustufe 2 – Ergänzende Quelladapter
  - Ausbaustufe 3 – Linked Data
  - Ausbaustufe 4 – KI
- ▶ Betrachtungen von Zeitrahmen, finanzielle und personelle Ressourcen sowie die Wirtschaftlichkeit betrachtet (siehe Kap. 4.5 in [1]);
- ▶ Eine prototypische Umsetzung des Webdesigns in Form eines Klick-Dummy (siehe Kap. 5 in [1]).

Die Machbarkeitsstudie kommt zu dem Ergebnis, dass umwelt.info aus technischer, organisatorischer und rechtlicher Sicht grundsätzlich umsetzbar ist. In der Zeitplanung wird ein Entwicklungszeitraum von etwa 5 Jahren vorgesehen. Es wird ein agiles Projektvorgehen vorgeschlagen, sodass bereits frühzeitig Zwischenergebnisse produktiv werden. Die Erarbeitung einer auf dieses Projekt angepassten agilen Vorgehensweise ist daher ein wesentlicher Bestandteil dieses Umsetzungskonzepts.

Das vorliegende Umsetzungskonzept umfasst eine Feinplanung der Umsetzung von umwelt.info und dient der Vorbereitung des agilen Prozesses für die technische Verwirklichung, der Ausarbeitung der technischen Konzeption sowie der Klärung organisatorischer und rechtlicher Fragestellungen:

- ▶ Ergänzung und Detaillierung der Zielgruppen und Bedarfe der Nutzenden (siehe Kap. 2)
- ▶ Technisches Konzept für eine Basisumsetzung und ergänzende Quelladapter des Portals umwelt.info – Ausbaustufen 1 und 2 der Machbarkeitsstudie (siehe Kap. 3 und Anhang A)
- ▶ Umsetzungskonzept für die Weiterentwicklung zu einem Linked Data-basierten und KI-unterstützten umwelt.info – Ausbaustufen 3 und 4 der Machbarkeitsstudie (siehe Kap. 4 und Anhang A)
- ▶ Rollen- und Vorgehensmodell für eine agile Implementierung der Basisumsetzung und ergänzender Quelladapter eines umwelt.info (siehe Kap. 5)
- ▶ Vorschläge für eine Vertragsgestaltung für eine agile Umsetzung (siehe Kap. 5.3)
- ▶ Konzept für die Organisation von Umsetzungsphase und nachfolgenden Betrieb (siehe Kap. 6)

Aktualisierte Betrachtung des Rechtsrahmens zur Umsetzung von umwelt.info (Kap. 7), Details zum Design sowie eine Aufgabenliste liegen in internen Dokumenten für die Entwicklung vor. Das Designkonzept enthält eine Darstellung der Anforderungen der Nutzenden (Personas, User Stories) sowie der Designentwürfe. Die Anforderungen basieren auf den Ergebnissen der Machbarkeitsstudie und wurden während der Bearbeitung des Umsetzungskonzepts in Teilen überarbeitet und ergänzt. Ein initialer Backlog mit einer Liste der Aufgaben für die Umsetzung wurde zunächst als Exceltabelle erstellt.

## 2 Ziele und Bedarfe der Nutzenden

In diesem Kapitel werden zunächst kurz Anlass und Ziele für umwelt.info und anschließend Nutzungsgruppen und deren Anwendungsfälle beschrieben.

### 2.1 Ziele von umwelt.info

Das Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV) formulierte in einer Pressemitteilung vom 26.08.2021 das Ziel: „Alle deutschen Umweltinformationen und -daten über eine Plattform erreichen – das ist das Ziel des neuen Nationalen Umweltinformationszentrums, ...“ [4]. Damit soll für mehr Transparenz und dadurch für eine verbesserte, innovative Nutzung von Umweltdaten und -informationen, gesorgt werden.

Als übergeordnete gesellschaftspolitische Ziele sind folgende Aspekte zu nennen (vgl. [1]):

- ▶ umwelt.info soll zur Erfüllung von EU-Richtlinien beitragen, insbesondere der Umweltinformationsrichtlinie (2003/4/EG), welche die Vorgaben der Aarhus-Konvention für das Recht auf freien Zugang zu Umweltinformationen auf EU-Ebene umsetzt.
- ▶ Im Kontext der „Strategie Künstliche Intelligenz“ des BMUV soll für Wissenschaft, Wirtschaft und Gesellschaft ein transparenter und valider Zugang zu Umweltdaten ermöglicht werden.
- ▶ Darüber hinaus ergeben sich unter anderem beispielhaft folgende konkrete Anwendungsfälle:
  - Beschleunigung in der Bauleitplanung durch einen gebündelten Zugang zu Umweltdaten für die Erstellung von Umweltberichten im Sinne des § 2 Abs. 4 BauGB.
  - Bereitstellung von Informationen zur Beurteilung der Gleichwertigkeit von Lebensverhältnissen.
  - Bereitstellung von Informationen über einen Schutzstatus von Gebieten und vorkommenden Tieren und Pflanzen im Zuge der Beschleunigung des Netzausbaus sowie weiterer Planungen.

Mit umwelt.info soll ein Umweltdatenportal entstehen, mit dem es möglich ist, an einer zentralen Stelle Daten und Informationen aus Deutschland zu unterschiedlichen umweltbezogenen Themenbereichen zu suchen und zu finden. Die Daten können strukturiert, semi-strukturiert oder unstrukturiert sein. Es sollen in umwelt.info aber keine Primärdaten gehalten werden. Somit wird umwelt.info Fachinformationssysteme, die auf allen Verwaltungsebenen existieren, nicht ersetzen. Insbesondere Fachsysteme der datenhaltenden Stellen haben weiterhin die Aufgabe, Daten einzelner Fachdomänen mit hoher Qualität bereitzustellen. Daher werden in umwelt.info Metadaten vorgehalten, sodass die Daten mit Informationen über die Daten verfügbar sind. Eine individuelle Anbindung ausgewählter Systeme oder die Weiterleitung auf Datenangebote kann für eine Auswahl von Datenbeständen Aufgabe von umwelt.info sein.

Weiterhin werden in umwelt.info redaktionelle Inhalte von berechtigten Personen erstellt, gepflegt und veröffentlicht. Redaktionelle Inhalte können z. B. Artikel, Videos, Infografiken oder Audioformate umfassen. Neben der Veröffentlichung auf dem Portal selbst dienen diese auch als Inhalte für soziale Netzwerke. Die redaktionellen Inhalte dienen insbesondere dazu, Informationen und Daten zielgruppen- und themengerecht aufzubereiten und für Bürgerinnen



und Bürger, Wissenschaft und Forschung, Unternehmen und Industrie sowie die Verwaltung zur Verfügung stellen [1].

Möglicher Nutzen für die datenhaltenden Stellen entsteht beispielsweise durch eine größere Reichweite der Daten. Datenhaltende Stellen sind in der Regel auch an der Nutzung von Daten interessiert, z. B. aus benachbarten oder vergleichbaren Regionen oder weiterer relevanter Fachthemen, die über umwelt.info auffindbar sind. Ähnliche Angebote anderer datenhaltender Stellen können auch als Vergleich und Ideengeber herangezogen werden.

Die nachfolgend dargestellte Produktvision stellt den Diskussionsstand zum Zeitpunkt der Erstellung dieses Berichtes dar und sollte am Beginn der Entwicklungsphase verifiziert und ggf. an aktuelle Rahmenbedingungen und Anforderungen angepasst werden. Dabei sollten möglichst viele Stakeholder des Projekts beteiligt werden. Nach dieser Verifizierung sollte die Vision nur in Ausnahmefällen verändert werden und als Richtgröße für die Prüfung der Projekt- und Sprintziele dienen. Der hier formulierte Entwurf ist direkt aus den Projektzielen abgeleitet.

#### **Vision für umwelt.info:**

- ▶ Ein Portal, über die alle deutschen Umwelteinformationen erreicht werden können;
- ▶ Ein Portal, das zu Umwelteinformationen aller Art Metadaten liefert, welche auf Qualität, Aktualität sowie Vollständigkeit prüfbar sind und Informationen zur Rechtssicherheit von Daten liefern;
- ▶ Ein System, das den Zugang zu Umweltdaten und -diensten gewährleistet;
- ▶ Ein Portal mit aggregierten und redaktionell aufbereiteten Informationen zu ausgewählten Umweltthemen;
- ▶ Ein Portal für eine heterogene Zielgruppe von Fachleuten bis hin zu interessierten Privatpersonen;
- ▶ Ein Portal, das Kommunikation und Zusammenarbeit fördert.

## **2.2 Nutzergruppen und User Stories**

Im Rahmen der Machbarkeitsstudie [1] wurde verschiedene Zielgruppen und deren Bedarfe mit Hilfe von Methoden aus dem Bereich des User-Centered Designs<sup>2</sup> analysiert. Neben Anforderungs-Workshops wurden Interviews mit Fachleuten durchgeführt. Die Analyse resultierte in 7 Personas<sup>3</sup>, 152 einzelnen User Stories, 17 User Story Clustern und 63 Best Practices, welche die Grundlage für das Designkonzept für umwelt.info bilden und in der Machbarkeitsstudie [1] dokumentiert sind.

Die Personas stellen eine Auswahl von umwelt.info-relevanten Nutzungsgruppen dar. Diese verdeutlichen, dass es sich bei den Nutzenden von umwelt.info um eine heterogene Zielgruppe handelt [1]:

- ▶ Mitarbeiter\*in im Datenmanagement in der öffentlichen Verwaltung arbeitet in einer Landesbehörde in der Datensammlung, Qualitätssicherung und Datenbereitstellung.

<sup>2</sup> Sehr viele Artikel zum Thema User Experience / User Centered Design findet man unter <https://www.nngroup.com/articles/>

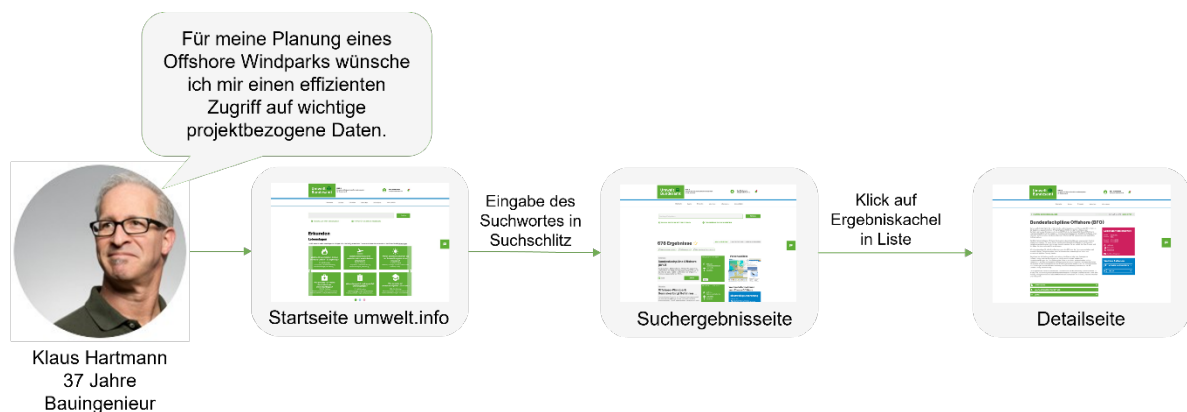
<sup>3</sup> Die Redakteurin (Persona 7), wurde nach der Machbarkeitsstudie ergänzt.

Gewünscht ist ein zentrales Metadatenportal mit einheitlichen Standards und hoher Datenqualität, auf dem Daten automatisiert veröffentlicht werden.

- ▶ Politikberater\*in auf lokaler Ebene wünscht sich einen einfachen Zugriff auf vielfältige, qualitätsgeprüfte und aktuelle Daten für die tägliche Arbeit.
- ▶ Beratende\*r aus dem Ingenieurwesen arbeitet an Infrastruktur-Großvorhaben und sucht für seine\*ihre Projekte regelmäßig und zielgerichtet nach validen Informationen.
- ▶ Umweltwissenschaftler\*in arbeitet an einer Universität oder anderen Forschungseinrichtungen. Das Portal soll die Datenrecherche durch umfangreiche Suchmöglichkeiten erleichtern und vollständige, valide und aufbereitete Metadaten an einem zentralen Ort liefern.
- ▶ Lehrer\*in wünscht sich für die Umweltbildung der Lernenden eine zentrale Plattform mit einer intelligenten und einfachen Suche, die durch zusammengestelltes Infomaterial zu Umweltthemen die Unterrichtsvorbereitung und Vermittlung erleichtert.
- ▶ Bürger\*in interessiert sich für Umweltthemen und wünscht sich einen zentralen Zugang und Benachrichtigungen zu Umweltinformationen aller Art (z. B. ein gesundes Leben mit umweltverträglicher Ressourcennutzung und nachhaltiger Lebensweise).
- ▶ Redakteur\*in kümmert sich um die redaktionelle Betreuung des Portals, das Verfassen von Texten und das Veröffentlichen von verschiedenen Medienarten. Er\*Sie wünscht sich ein flexibles CMS, mit dem neue Ideen für Formate und Formen der Veröffentlichung ausgetestet werden können.

Die unten genannten User Journeys (Abbildung 1 bis Abbildung 4) stellen wichtige Nutzungsszenarien von 4 Personas dar und helfen sicherzustellen, dass die unterschiedlichen Anforderungen beachtet werden und die Nutzenden intuitiv zu dem Ziel ihrer Suche finden.

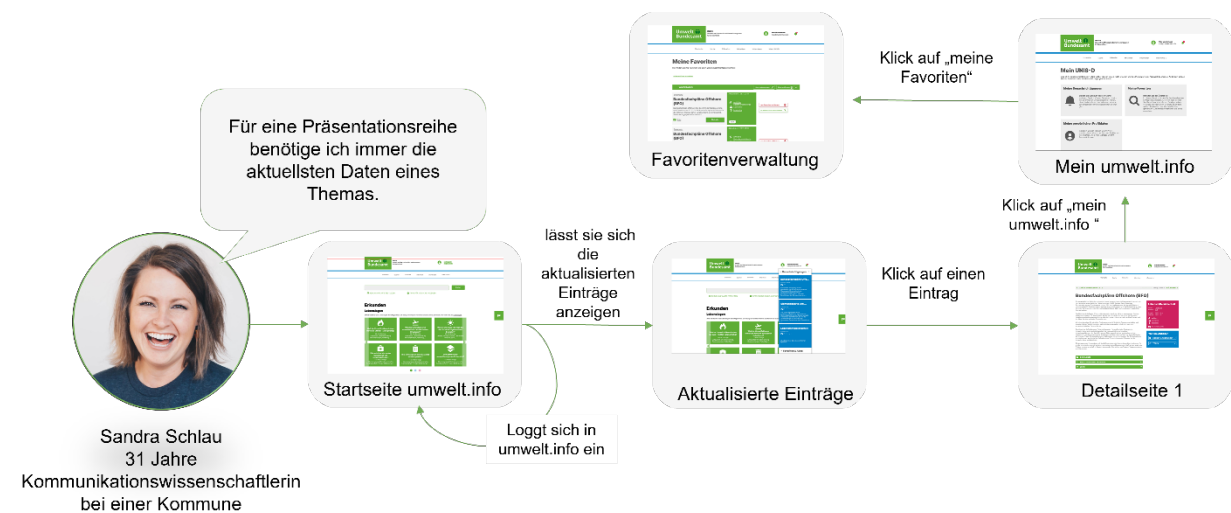
**Abbildung 1: User Journey Map: Beratende Ingenieur\*in**



Quelle: eigene Darstellung, con terra

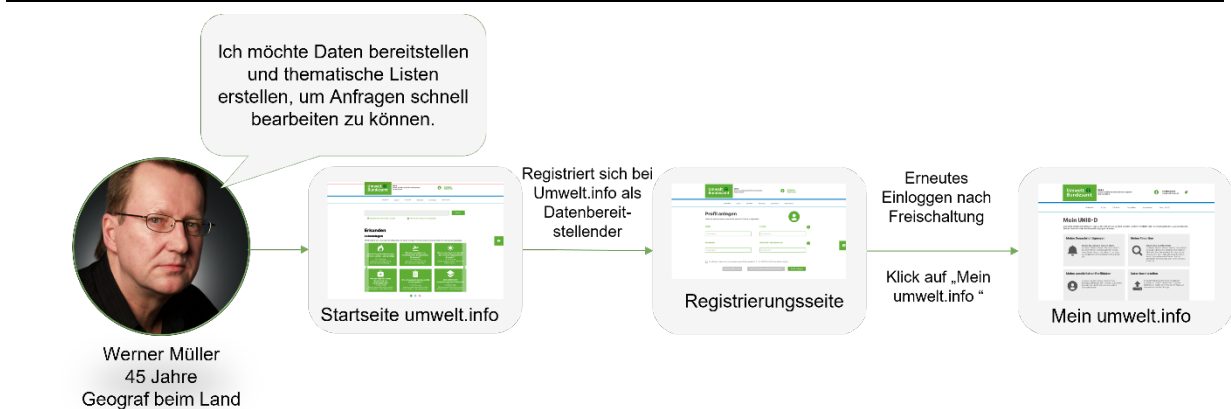


**Abbildung 2: User Journey Map: Politikberater\*in auf lokaler Ebene**



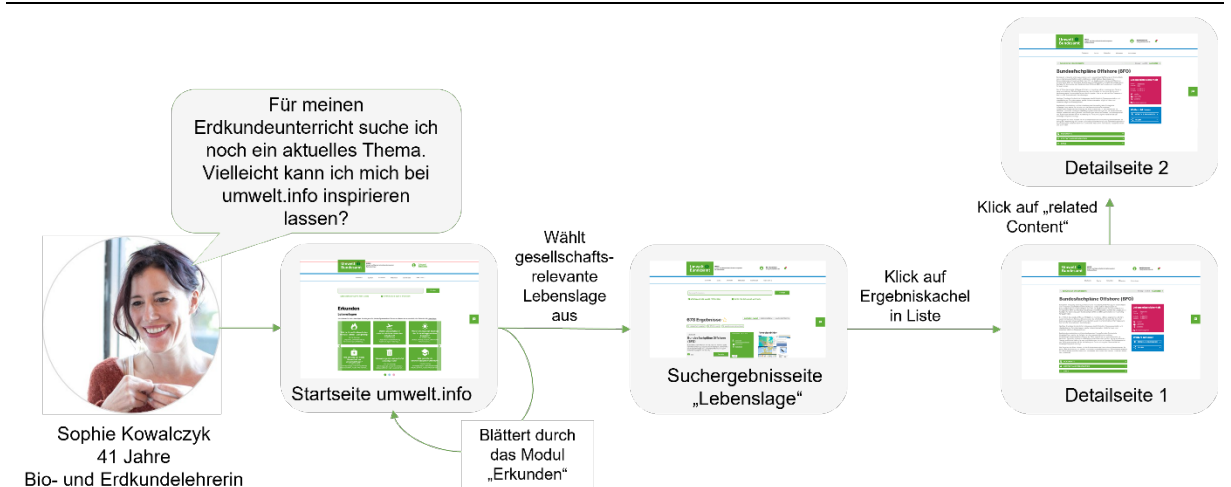
Quelle: eigene Darstellung, con terra

**Abbildung 3: User Journey Map: Mitarbeiter\*in Datenmanagement öffentliche Verwaltung**



Quelle: eigene Darstellung, con terra

**Abbildung 4: User Journey Map: Lehrer\*in**



Quelle: eigene Darstellung, con terra

Zusammenfassend zeigte die Analyse der Zielgruppen und deren Bedarfe in der Machbarkeitsstudie [1] einen großen Bedarf an einem zentralen Portal in der zersplitterten Landschaft der Umweltinformationen in Deutschland. Die Förderung der Anwendung von Standards für Metadaten und Schnittstellen sowie die Erfüllung von Qualitätsanforderungen (Aktualität, Vollständigkeit, Rechtssicherheit, Validität und Maschinenlesbarkeit) sind wesentliche Ziele für umwelt.info. Diese Ziele sollen bereits im Vorfeld der Datenbereitstellung sowie bei der Integration in das zentrale Umweltportal verfolgt werden. Den unterschiedlichen Bedürfnissen einer großen heterogenen Zielgruppe kann umwelt.info durch eine Vielfalt an Informationsformaten und Daten gerecht werden (vgl. [1]).

Auf Basis der Anforderungen der Personas wurde ein Design-Konzept erstellt, das als internes Dokument im Umweltbundesamt vorliegt. Darin sind die Oberflächenentwürfe des Systems inklusive erläuternder Design-Hinweise umfassend dargestellt. Die wesentlichen Oberflächenentwürfe sind bereits in der Machbarkeitsstudie enthalten [1].

### 3 Systemarchitektur

In diesem Kapitel wird die Systemarchitektur beschrieben. Die Architektur basiert auf den Ergebnissen der Machbarkeitsstudie [1], technischen Randbedingungen, Qualitätszielen und den durchgeführten Workshops zu den Themen „Linked Data“ und „Künstliche Intelligenz“ (KI). Es werden offene Fragen aus der Machbarkeitsstudie beantwortet bzw. bestehende Lücken vollständig konzipiert. Dieses Kapitel, in Verbindung mit Anhang A, stellt somit eine IT-Konzeption als Basis für die anschließende Umsetzung dar. Eine detaillierte Beschreibung aller Komponenten und Subkomponenten ist dem Anhang A zu entnehmen. Methodisch und inhaltlich basiert die Architekturdokumentation auf dem Arc42 Architektur Template [5], das eine Vorlage zur Entwicklung, Dokumentation und Kommunikation von Software- und Systemarchitekturen bietet. Wesentliche Zielgruppen der Beschreibung der Systemarchitektur sind Personen, die an der Umsetzung von umwelt.info arbeiten oder in Entscheidungen über das System und dessen Entwicklung beteiligt sind. Dies sind insbesondere das Entwicklungsteam, die Projektleitung im Umweltbundesamt, und das Betriebsteam.

#### 3.1 Technische Randbedingungen und Qualitätsziele

In diesem Kapitel werden die technische Randbedingungen für die Systemarchitektur beschrieben. Diese orientieren sich an den Architekturrichtlinien der IT des Bundes [6] und sind für umwelt.info angepasst in der nachfolgend aufgeführten Tabelle 1 aufgelistet.

**Tabelle 1: Technische Randbedingungen der IT des Bundes**

Bezeichnung	Vorgaben	Verbindlichkeitsgrad <sup>4</sup>
Content Management System	Government Site Builder	SOLL
Webserver	Apache HTTP Server Nginx	SOLL
API-Managementsysteme	Open API Specification 3	SOLL
Relationale Datenbankmanagementsysteme	MariaDB Microsoft SQL Server Oracle Database Oracle MySQL PostgreSQL	SOLL
Schnittstellen zur Kommunikation mit SQL-Datenbanken	Java: Java Persistence API / Jakarta Persistence Query Language PHP: PHP Data Objects	SOLL
Strukturierung von Webseiten	Hypertext Markup Language 5	SOLL
Layout und das Design von Webseiten	Cascading Style Sheets, Level 2 & 3	SOLL
Softwareentwicklungsumgebung	Java-Anwendungen: Eclipse, JetBrains IntelliJ IDEA PHP: Eclipse, JetBrains WebStorm Python: Eclipse, JetBrains PyCharm	SOLL

<sup>4</sup> Nach [6]: „MUSS“ kennzeichnet eine Aussage mit dem Charakter einer verbindlichen Festlegung. „SOLL“ kennzeichnet eine verbindliche Aussage, von der bei Vorliegen wesentlicher Gründe abgewichen werden kann. Die aus der Abweichung resultierenden Auswirkungen sind sorgfältig abzuwägen. Die Abweichungen sind zu dokumentieren.

Bezeichnung	Vorgaben	Verbindlichkeitsgrad <sup>4</sup>
	.NET-Anwendungen (C#): Eclipse (aCute), Microsoft Visual Studio	
Programmiersprachen (Empfehlung LTS-Version)	Java PHP C++ C# Python ABAP* (nur als Ergänzung für SAP- Anwendungen)	SOLL
Betriebssysteme	Microsoft Windows Server 2019 Red Hat Enterprise Linux 8 (RHEL) SUSE Linux Enterprise Server 15 (SLES)	SOLL
Kodierrichtlinien	Für die Kodierung von Zeichensätzen in den Software-Systemen soll das UCS Transformation Format (UTF-8) verwendet werden. Bei der Datenverarbeitung (speziell Metadatenverarbeitung) soll das ASCII Format zum Einsatz kommen.	SOLL
Versionsverwaltungssysteme	GIT	SOLL
Technische Plattform	Rechenzentrum (intern im Umweltbundesamt) Cloudinfrastruktur (z. B. ITZ- Bund/„Bundescloud“)	SOLL
Security - Übertragung von Daten zwischen Client und Webserver	Transport Layer Security	SOLL
Security - Protokolle für Authentifizierung und Autorisierung	Lightweight Directory Access Protocol, Version 3 (LDAP) Security Assertion Markup Language 2.0 OAuth 2.0 / OIDC (OpenID Connect)	SOLL
Kommunikation - Mail-Servern	Simple Mail Transfer Protocol Safe	MUSS
Kommunikation - E-Mail-Verkehrs	Multipurpose Internet Mail Extensions	MUSS
Kommunikation - Übertragung elektronischer Daten	Online Service Computer Interface- Transport 1.2	MUSS
Containerisierung	Docker Podman Kubernetes	SOLL
Laufzeitumgebungen und Frameworks	Java: Java Enterprise Edition ab Version 8 (Jakarta   Java EE), Spring ab Version 5 Microsoft .NET: Microsoft .NET Standard, Microsoft .NET Core	SOLL
Frameworks für die Webentwicklung	PHP: Symfony, Laravel Python: Django, TurboGears, Web2py	KANN
Anwendungsserver	Java Servlet/JSP-Container: Apache Tomcat, Undertow	SOLL

Bezeichnung	Vorgaben	Verbindlichkeitsgrad <sup>4</sup>
	Java EE-Applikationsserver: WildFly Application Server, Red Hat JBoss EAP, IBM WebSphere Application Server (WAS) PHP / Python / Ruby Webserver: Apache HTTP Server, Nginx .NET kompatible Webserver: Microsoft Internet Information Services (IIS), Kestrel (ASP.NET Core)	
kontinuierliche Integration	Jenkins GitLab CI	SOLL
Bereitstellung von Diensten	Representational State Transfer (REST) Simple Object Access Protocol	SOLL
OGC-konforme Dienste	Catalogue Services Specification 2.02 Web Map Service 1.3 (WMS) Web Coverage Service 2.1 Web Feature Service 2	MUSS
Strukturierten Daten zum Zweck des Datenaustausches	Extensible Markup Language (XML) JavaScript Object Notation (JSON) YAML	MUSS
Beschreibungssprachen für Metadaten	Resource Description Framework 1.1 (RDF) Dublin Core Metadata Element Set	SOLL

Bei Entscheidungen und Empfehlungen werden diese Randbedingungen berücksichtigt. Empfehlungen für bestimmte Architekturmuster, Technologien und Komponenten werden im Anhang A behandelt. Dabei werden die wichtigen Entscheidungsmerkmale inklusive der jeweiligen Begründung für die Auswahl aufgeführt.

Berücksichtigt wurden auch die in Tabelle 2 genannten wesentlichen Qualitätsziele.

**Tabelle 2: Qualitätsziele**

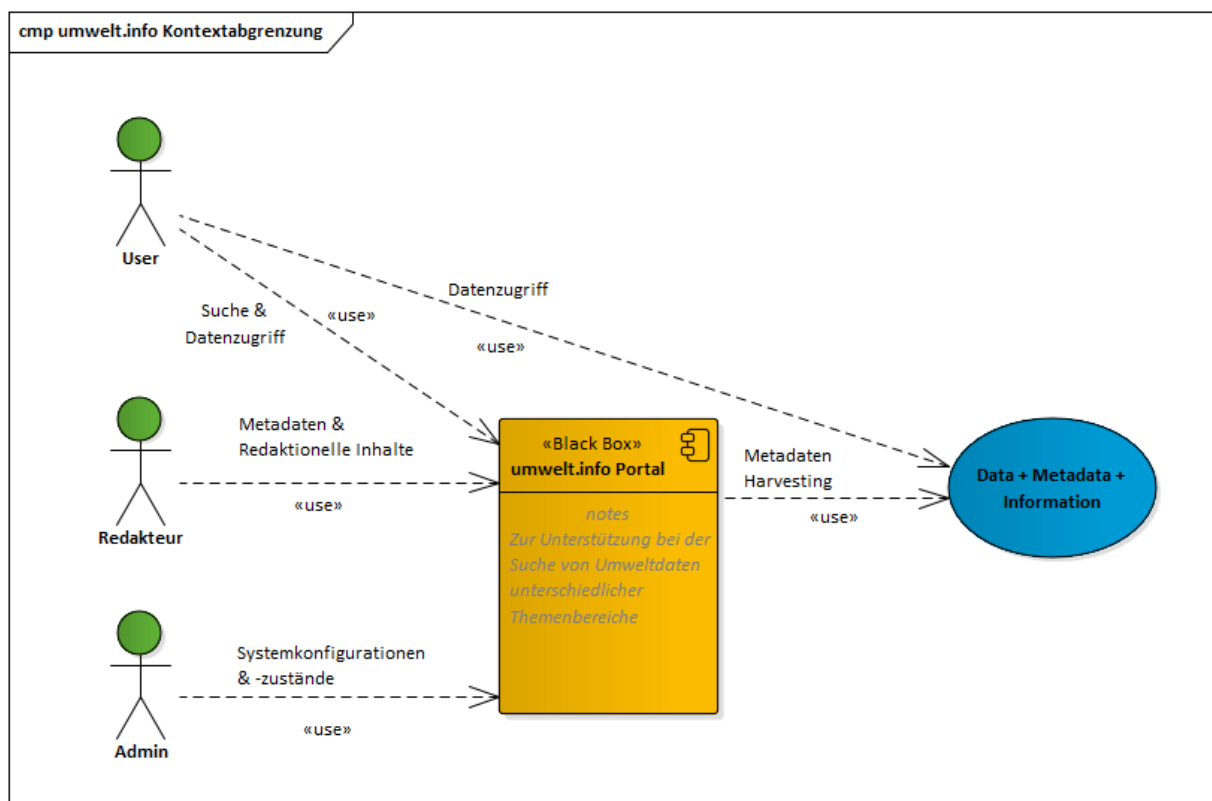
Qualitätsmerkmal	Motivation und Erläuterung
(Meta-)Datenqualität	<p>Die (Meta-)Datenqualität (z. B. Aktualität von Datensätzen, Zusammenfassung/Dopplung einzelner Datensätze, Formulierung und Formatierung der Beschreibungstexte, Kontaktpersonen und Ansprechpartner, etc.) hat einen großen Einfluss auf die Nutzbarkeit und Akzeptanz von umwelt.info als Informationssystem (siehe Machbarkeitsstudie). Deshalb sollten nicht nur Datenbereitsteller angehalten werden, entsprechend qualitativ hochwertige (Meta-)Daten bzw. Inhalte bereitzustellen, sondern auch der Harvesting-/Crawling-Prozess sollte hier seinen Beitrag leisten.</p> <p>Zu den Qualitätsanforderungen an den Harvesting-/Crawling-Prozess gehören u.a.:</p> <ul style="list-style-type: none"> <li>► Sicherstellen, dass die Metadaten, die in das System übernommen werden, die Anforderungen des minimalen gemeinsamen Metadatenmodells (Basis-Metadaten) erfüllen</li> </ul>

Qualitätsmerkmal	Motivation und Erläuterung
	<ul style="list-style-type: none"> <li>► Sofern für bestimmte textliche Informationen (z. B. Schlüsselworte, Organisationsbeschreibungen, Formattypen, Zugriffsverfahren) nicht bereits die Begrifflichkeiten gemeinsamer Thesauri verwendet werden, sollte ein „Mapping“ der Informationen auf Begrifflichkeiten/Objekte der gemeinsamen Thesauri versucht werden</li> <li>• Ein Geo-Thesaurus „Mapping“ sollte versucht werden, um aus den (Meta-)Daten intelligent räumliche Beschreibungen (inklusive Koordinaten der räumlichen Ausdehnung) abzuleiten</li> <li>► Idealerweise sollte eine Voransicht („Preview“) aus den Daten abgeleitet werden</li> </ul> <p>Ebenso ist es aber auch wichtig, dass neue oder geänderte Metadaten schnell in das System übernommen werden.</p>
Vereinfachte Aktualisierbarkeit / Testbarkeit	Es soll eine vereinfachte Aktualisier- und Testbarkeit des Systems durch eine Aufteilung in verschiedene Funktionen erreicht werden. Es ist wesentlich einfacher und sicherer, einzelne Funktionen des Systems zu bearbeiten, zu aktualisieren, zu testen und wiederzuverwenden.
Vereinfachte und automatisierte Bereitstellung neu entwickelter oder geänderter Funktionen	Es soll eine vereinfachte, schnelle und automatisierte Bereitstellung neu entwickelter oder geänderter Funktionen in das System möglich sein (s. auch „DevOps“ Software Entwicklungspraktiken).
Vereinfachte Skalierbarkeit des Gesamtsystems in verschiedene Umgebungen	umwelt.info soll so entwickelt werden, dass es sowohl in einer („Public“) „Cloud“-Infrastruktur als auch auf dedizierten Virtuellen Maschinen (z. B. in einem Rechenzentrum) betrieben werden kann. Dadurch kann das System unabhängig von der Umgebung entwickelt werden, ist auf unterschiedlichen Plattformen installier- und ausführbar und bietet verschiedene Mechanismen für eine Skalierung.
Einsatz etablierter Technologien für eine erhöhte Sicherheit, Verfügbarkeit und Zuverlässigkeit	Bei der Entwicklung des umwelt.info Gesamtsystems soll der Einsatz etablierter Technologien favorisiert werden. Damit soll das System vor allem stabil und getestet sein und so besseren Schutz vor Softwareangriffen und Ausfall standhalten. Ein weiterer wesentlicher Punkt ist, die breite Wissensbasis für die Entwicklung, die potenzielle Weiterentwicklung und Nachhaltigkeit.

## 3.2 Kontextabgrenzung

In der folgenden Abbildung ist eine grobe Kontextabgrenzung des umwelt.info Portals dargestellt. Diese Darstellung zeigt in einer Top-Level-Sicht das Gesamtsystem des Portals und dessen Beziehungen zur Außenwelt (verändert nach [1]).

**Abbildung 5: Kontextabgrenzung des umwelt.info-Portals**



Quelle: eigene Darstellung, con terra GmbH

Die Tabelle 3 zeigt die Rollen im umwelt.info Portal mit ihren Kommunikationsbeziehungen.

**Tabelle 3: Kommunikationsbeziehungen des umwelt.info Portals (Fachlicher Kontext)**

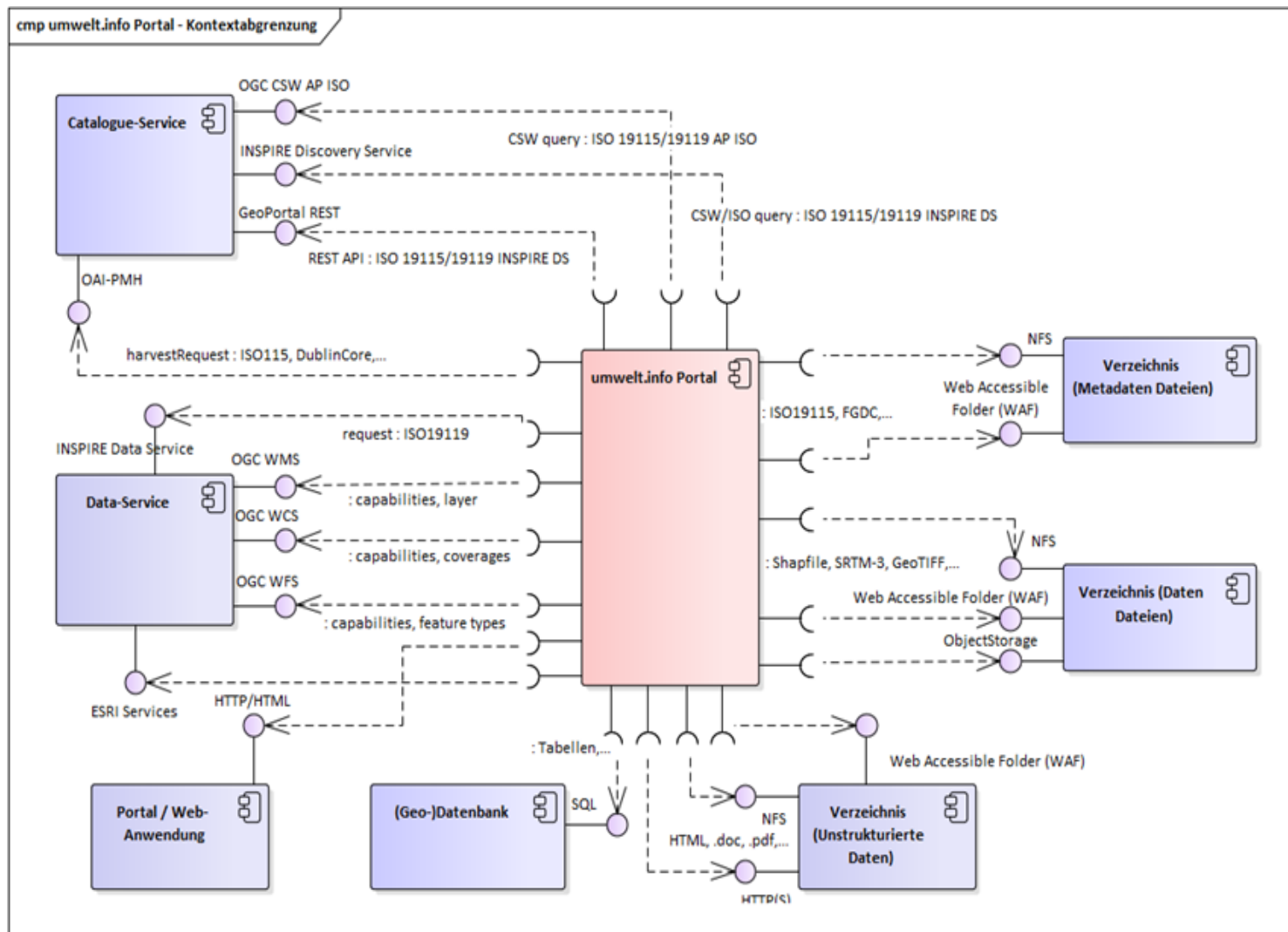
Kommunikationsbeziehung	Eingabe	Ausgabe
<b>Nutzende</b> (User) suchen nach Umweltdaten und -informationen im umwelt.info Portal (zu unterschiedlichen Themenbereichen) und möchten bei Bedarf über die Metadaten auf die Daten zugreifen; optional gibt es für Nutzende eine Registrierung mit einigen zusätzlichen Funktionen	Suchanfrage: Suchkriterien (Thema, Datentyp, Raumbezug, etc.) Datenzugriffsanfrage: Pfad, Uniform Resource Locator (URL), Parameter, Format, etc.	Metadaten zu den Umweltdaten und -informationen (idealerweise mit enthaltenen Informationen, um über geeignete Schnittstellen auf diese zugreifen zu können) Daten und Informationen
<b>Administrierende</b> überwachen und steuern den laufenden Betrieb des <b>umwelt.info Portals</b>	Konfigurationsparameter, Parameter für die Abfrage von Laufzeitinformationen, etc.  Neue Datenquelle	Systemparameter, Infos über erfolgreiche/nicht erfolgreiche Änderungen am Systemzustand, etc.  Bestätigung der Einbindung der neuen Datenquelle
<b>CMS-Redakteur*innen</b> verwalten und pflegen die redaktionellen Inhalte des <b>umwelt.info Portals</b>	Redaktionelle Inhalte	Bestätigung über erfolgreiche / nicht erfolgreiche Erstellung / Veröffentlichung von Inhalten

Kommunikationsbeziehung	Eingabe	Ausgabe
<b>Metadaten-Redakteur*innen</b> verwalten und pflegen die Metadaten des <b>umwelt.info Portals</b>	Metadaten zu Daten, Informationen und Anwendungen, Beschreibungen, etc.	Bestätigung über erfolgreiche / nicht erfolgreiche Änderung der Metadaten, Beschreibungen, etc.
<b>umwelt.info</b> „harvested“ Metadaten und Zugriffsinformationen zu außerhalb des Portals liegenden <b>Datenquellen</b> und speichert und indiziert diese Metadaten (Metadaten-Index) und Zugriffsinformationen	Harvesting-Anfrage: Filter- und Iterationskriterien	Metadaten zu den Umweltdaten und -informationen (evtl. mit Informationen, um über geeignete Schnittstellen diese zugreifen zu können)

In der Abbildung 6 ist das Portal mit seinen Außenbeziehungen aus technischer Sicht dargestellt. Je nach zu indizierender Datenquelle sind für das Harvesting der Metadaten unterschiedliche Zugriffsmechanismen (aufgrund unterschiedlicher Schnittstellen) nötig. Demnach werden unterschiedliche Metadaten bzw. Metadaten-Encodings als Ergebnis geliefert.



**Abbildung 6: Kontextabgrenzung (technisch) des umwelt.info-Umweltportals**



Quelle: eigene Darstellung, con terra GmbH

Der Zugriff auf die (Meta-) Daten lässt sich entlang mehrerer Dimensionen typisieren und folgendermaßen klassifizieren:

**Tabelle 4: Input und Output des umwelt.info Portals (Technischer Kontext) (verändert nach [1])**

Typ der Datenquelle	Beschreibung	Input	Output
(Standardisierte) web-basierte Metadaten-Dienste mit (standardisiertem) Metadaten-Profil.	Zu den wichtigsten standardisierten Metadaten-Diensten mit bekanntem Metadatenprofil und Zugriffsprotokoll gehören etwa das OGC Catalogue Service Web (CSW) ISO Application Profile (AP) ISO, OGC API Records, INSPIRE Discovery Services, das "Open Archives Initiative - Protocol for Metadata Harvesting" interface oder etwa die ESRI GeoPortal API.	HTTP GET/KVP Request  HTTP POST/XML Request	ISO19115/19119 (INSPIRE DS) XML Dublin Core XML/JSON
(Standardisierte) web-basierte Daten-Dienste (mit standardisiertem Daten-Profil) aus denen Metadaten noch zu extrahieren sind.	Zu den wichtigsten standardisierten Daten-Diensten mit bekanntem Service-Metadaten-Profil (Capabilities Dokumente) oder Datenprofil gehören OGC Web Feature Services, OGC Web Map Services (WMS), OGC Web Coverage Services, OGC API Features/Coverages/..., ESRI Daten-Dienste (Map-Services, Geodatabases, etc.)	HTTP(S) GET/KVP Request  HTTP(S) POST/XML Request	OGC Capabilities XML OpenAPI JSON ...
Dateien in Netzwerk- oder Web-Verzeichnissen mit bekannten Metadatenprofilen.	Zu den wichtigsten Dateien in Netzwerk- oder Web-Verzeichnissen mit bekanntem Metadatenprofil gehören ISO19115- oder Dublin Core XML Files, GeoPackages und spezielle Esri-Formate	NFS Request  Web Accessible Folder (WAF) Request	ISO19115/19119 (INSPIRE DS) XML  Dublin Core XML/JSON
Dateien in Netzwerk- oder Web-Verzeichnissen mit bekannten Datenprofilen.	Zu den wichtigsten Dateien in Netzwerk- oder Web-Verzeichnissen mit bekanntem Datenprofil gehören bekannte Rasterdateien (z. B. GeoTIFF), Geländemodelle (z. B. SRTM-3), Shape-Files, ESA SAFE Files, Excel-Files, Statistiken mit bekanntem Format.	NFS Request  Web Accessible Folder (WAF) Request	GeoTIFF), Geländemodelle (z. B. SRTM-3), Shape-Files, ESA SAFE Files, Excel-Files, Statistiken mit bekanntem Format.
(Geo-)Datenbanken: Datenbanken mit oder ohne Geoinformation (z. B. Stoffdatenbanken, Forschungsdatenbanken)			

Typ der Datenquelle	Beschreibung	Input	Output
Dateien in Netzwerk- oder Web-Verzeichnissen mit unstrukturierten Daten und Informationen (Web-Seiten, Word- oder PDF-Dokumente, etc.)	Berichte, Studien, Informationen zu Öffentlichkeitsarbeit, unstrukturierte Statistiken, Präsentationen, Bilder, Daten-Produkte	HTTP(S) NFS Request Web Accessible Folder (WAF) Request	Word- oder PDF-Dokumente PowerPoint HTML Bilder Excel
Portale / Web-Anwendungen	Anwendungen mit Umweltdaten, bei denen die Inhalte nicht automatisch indiziert werden können	HTTP(S)	HTML

### 3.3 Lösungsstrategie

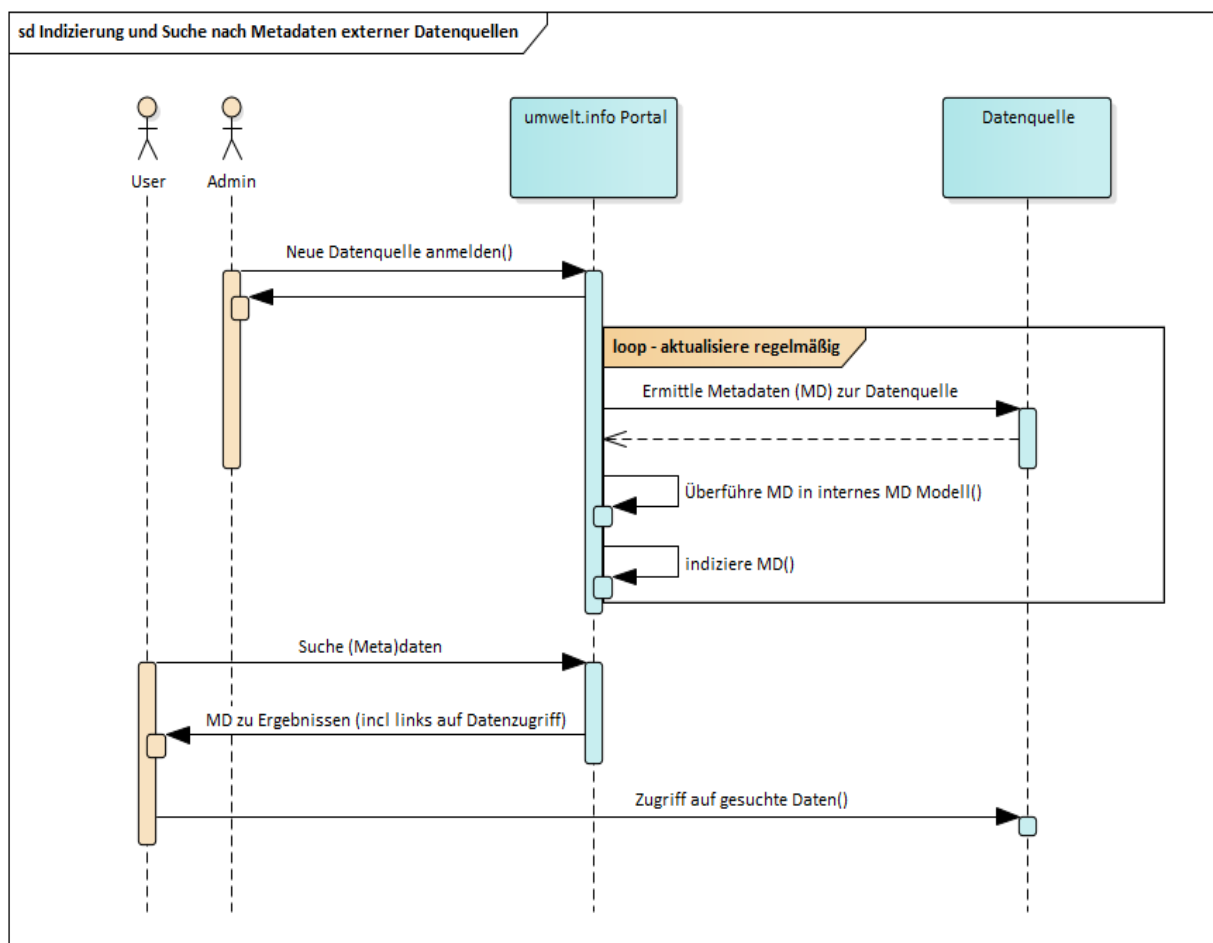
Auf Basis der User Stories (siehe [1]) wurden größere Funktionsbereiche herausgearbeitet, die in der Scrum Methode auch als Epic bezeichnet werden. Diese werden in den nächsten Abschnitten erläutert. Den Themen Linked Data und Künstliche Intelligenz wird ein eigenes Kapitel gewidmet (siehe Kap. 4).

#### 3.3.1 Harvesting und Crawling von Metadaten

Für eine effiziente Suche nach Daten und Informationen ist der Aufbau eines Suchindex erforderlich. Die Inhalte, die später in umwelt.info gefunden werden sollen, müssen dazu für diesen Index aufbereitet werden. Für diesen Zweck werden Harvester und Crawler vorgesehen.

Das Metadaten-Harvesting überführt bereits bestehende Metadaten aus externen Quellen in den Metadaten-Index von umwelt.info. Crawler hingegen suchen Inhalte im Internet und generieren aus diesen die Indexeinträge. Das UML-Sequenzdiagramm in Abbildung 7 zeigt das Zusammenspiel der Komponenten, die für das Bekanntmachen und Indizieren der Daten und Informationen sowie die Suche nach verfügbaren Daten und Informationen notwendig sind.

**Abbildung 7: Genereller Ablauf beim Harvesting und Crawling**



Quelle: eigene Darstellung, con terra GmbH

Für jede Art einer externen Datenquelle existiert ein spezieller „Adaptor“. Damit findet sich die Komplexität der zu indizierenden Datenquellen ausschließlich in den Adaptoren. Dieser Lösungsansatz wurde auch schon im Europäischen Datenportal (EDP) verwendet [7]. Die Adaptoren müssen in der Lage sein, bestimmte Filterungen der (Meta-)daten durchzuführen, die nutzbaren Metadaten zu extrahieren und semantische Tests vorzunehmen. Das Metadatenmodell der im Index gespeicherten Metadaten setzt dabei auf weitverbreitete und mächtige Standards auf, die auch „Linked Data tauglich“ sind: DCAT-AP und Schema.org.

Der beschriebene Lösungsansatz kann für jede Art von externer Datenquelle verwendet werden. Auch für die Integration und Nutzbarkeit großer Dateninfrastrukturen, wie der Nationalen Forschungsdateninfrastruktur, der Marinen Dateninfrastruktur, sowie dem Zentralen Informationsknoten Wasserwirtschaft Deutschland (Wasser-DE), ist dieser Lösungsansatz einsetzbar. Es müssen dafür jeweils eigene Harvester oder Crawler bereitgestellt werden, die die Metadaten (ohne Doppelungen) in das umwelt.info System integrieren.

### 3.3.2 Suche

Das Portal umwelt.info ermöglicht den Nutzenden eine zentrale Suche für sämtliche Umwelt- und Naturschutzthemen. Die Suche muss einfach und intelligent durchführbar sein und einen niedrigschwelligen, möglichst direkten Zugang zu den Metadaten (und somit allen referenzierten Informationsquellen) ermöglichen. Vorgesehen sind Oberflächen für die Suche, die Suchergebnisliste und die Detailansicht. Für die Suche gibt es eine Volltextsuche, Filter und

räumliche Eingrenzungen. Die verschiedenen Suchfunktionen des Portals werden in den CMS-Seiten entsprechend eingebunden.

Die Suchergebnisliste repräsentiert Metadaten, die in der Detailansicht tiefergehend betrachtet und heruntergeladen werden können. Wichtige Attribute (wie beispielsweise Datenbereitsteller, Lizenz und Kosten) sind bereits in der Suchergebnisliste ersichtlich. Um einen einfachen und direkten Zugriff auf die Daten selbst zu ermöglichen, enthält die Metadatenrepräsentation einen Link auf die eigentlichen Daten. Für angemeldete Nutzende werden auf diesen Seiten zudem die Profilkfunktionen aktiv (vgl. Kap. 3.3.7).

Für die Suche werden die Metadaten im Metadatenindex in einem speziellen Suchschema gespeichert. Um eine effiziente und effektive Suche auf dem Metadaten-Index zu ermöglichen, müssen in der Benutzeroberfläche Funktionen wie Volltextsuche, Filter und räumliche Eingrenzung bereitgestellt werden. Diese verschiedenen Suchfunktionen werden in die CMS-Seiten entsprechend eingebunden und vom Metadatenindex serverseitig unterstützt. Der Metadatenindex bedient viele Funktionen mit Standardfunktionalität (etwa Textindexierung oder spezielle Felder für Facetten). So ist auch eine Option für die Sortierung der Suchergebnisse nach Relevanz enthalten. Hierzu kann zunächst der Standardalgorithmus der Suchmaschine genutzt werden.

Während der Eingabe einer Suche in den Suchschlitz erhalten die Nutzenden bereits Treffer zur Auswahl angezeigt. Dabei werden bereits während der Eingabe im Hintergrund Suchanfragen an den Metadaten-Index gestellt, um für die unterschiedlichen Gruppen (z. B. Ressource-Typen) die besten Treffer bereits in einer (mit dem Suchschlitz integrierten) Auswahlliste anzuzeigen.

Durch die Suche ergeben sich Anforderungen an das Metadatenmodell und andere Komponenten von umwelt.info. Insbesondere werden für Facetten und Attribute in der Benutzeroberfläche entsprechende Felder im Metadatenschema benötigt. Für individuelle Einstellungen oder gespeicherte Suchen wird ein Nutzerprofil benötigt und die Qualität der (Meta-)Daten muss beim Harvesting und Editieren erfasst werden.

### **3.3.3 Redaktionelle Inhalte einstellen**

Neben den Metadaten sollen auch redaktionelle Inhalte im Portal verfügbar sein. Diese werden von Redakteur\*innen erstellt und gepflegt, wobei verschiedene Medienarten eingebunden werden können (z. B. Texte, Bilder, Videos, Audioformate oder externe Inhalte). Die verschiedenen redaktionellen Inhalte werden auf der Startseite angezeigt und bieten so einen alternativen Einstiegspunkt zur Suche. Redaktionelle Inhalte als Einstiegspunkte bieten die Möglichkeit bestimmte Daten und Informationsquellen hervorzuheben und sind auf der Startseite des Portals als „Erkunden“, „Highlights“ und „Aktuelles“ sichtbar. Zum Beispiel kann auf einzelne Datensätze, die im Metadatenindex von umwelt.info eingetragen wurden, aufmerksam gemacht werden. Die Sektion „Erkunden“ umfasst die drei Kategorien „Lebenslagen“, „Fachkategorien“ und „Zielgruppen“. Die genannten Kategorien führen jeweils auf eine eigene Seite mit redaktionellen Inhalten. Die dort verlinkten Seiten zeigen neben dem eigentlichen Seiteninhalt auch redaktionell erstellte Listen von Datensätzen aus dem Metadatenportal.

Die Umsetzung wird im Wesentlichen durch die Komponente „Content Management System (CMS)“ realisiert, die den Redakteur\*innen eine geeignete Benutzeroberfläche für die Erstellung, Pflege und Verwaltung der Seiten des Portals bietet. Darüber hinaus können Nutzungsoberflächen des umwelt.info-Portals (z. B. Suchergebnislisten oder Detailansichten) in die mit dem CMS erstellten Seiten eingebettet werden. Um Nutzeroberflächen (z. B. Suchergebnislisten oder Detailansichten) in der Seite zu platzieren werden in Hypertext Markup

Language (HTML) Platzhalterelemente definiert, die von den Redakteur\*innen in ihren Inhalten eingesetzt und konfiguriert werden können.

### **3.3.4 Metadaten manuell erfassen oder importieren**

„Metadaten-Redakteur\*innen“ erhalten die Möglichkeit, über den „Data Check-In“ Metadaten für Datenquellen, für die keine Möglichkeit des Harvesting besteht, manuell zu erfassen oder zu importieren. Zu diesem Zweck wird ein Web-basierter Metadaten-Editor bereitgestellt. Mit dieser Komponente können Metadaten erfasst werden, die dem internen Metadatenmodell des umwelt.info-Portals entsprechen. Neben den „Standardmetadaten“ lassen sich damit auch die Informationen zum Zugriff auf die beschriebene Datenquelle erfassen. Der Import bereits existierender Metadaten (z. B. einer Datei aus dem Dateiverzeichnis oder über eine URL) und die Nachbearbeitung oder das Löschen von bestehenden Metadatensätzen ist ebenso möglich. Die Erfassung der räumlichen Ausdehnung wird gemäß DCAT-AP (Kreis, Bezirk, etc.) angegeben und vom Metadata-Editor als geografische Ausdehnung im Metadaten-Index gespeichert. Es soll zudem möglich sein über einen automatisierten Prozess (welcher noch näher in der Umsetzung zu definieren ist), beispielsweise über ein Formular, Metadaten von externen Datenbereitstellenden durch Metadaten-Redakteur\*innen in umwelt.info zu übernehmen. Für die Nutzung dieser Funktion wird eine entsprechende Registrierung am umwelt.info Portal benötigt.

### **3.3.5 Daten nutzen / neue Daten ableiten**

Im umwelt.info-Portal sollen den Nutzenden Anwendungen bereitgestellt werden, mit denen Umweltdaten visualisiert und neue Erkenntnisse gewonnen werden können. Dabei kann es sich um unterschiedliche Datentypen handeln, wie z. B. Geodaten oder Zeitreihen. Die Anforderungen an die Datennutzung sind noch nicht vollständig beschrieben. Welche Anwendungen es geben wird, ist im weiteren Projektverlauf noch zu definieren. In umwelt.info soll daher eine erweiterbare Komponente „Application“ integriert werden. Dadurch besteht die Möglichkeit, weitere Daten-Anwendungen für die Nutzenden stetig zu evaluieren und in die „Application“ von umwelt.info zu implementieren. Die „Application“ erkennt, um welche Datentypen („mimeType“) es sich handelt, sodass automatisch die dazu passende Anwendung gestartet wird (so wie etwa Windows erkennt, dass der PDF-Reader zu starten ist, wenn eine pdf-Datei geöffnet werden soll). Zum Beispiel würde sich eine Kartenanwendung öffnen, wenn die Nutzenden einen WMS-Dienst gefunden haben. In der Anwendung können auch eigene Daten hinzugeladen werden, sofern diese Funktionalität von der jeweiligen Anwendung bereitgestellt wird. Auch eine Integration von Funktionen des beim Umweltbundesamt in Entwicklung befindlichen „Data Cube“ ist an dieser Stelle denkbar.

### **3.3.6 Nutzerverwaltung**

Das umwelt.info-Portal stellt auch Funktionen bereit, die nur für angemeldete Nutzende bzw. umwelt.info-Mitarbeitende verfügbar sind, wie beispielsweise das Einstellen redaktioneller Inhalte oder die Publikation von Metadaten. Daher verfügt umwelt.info über eine eigene Nutzerverwaltung. Die darin registrierten Nutzenden können sich über eine Authentifizierungskomponente am Portal anmelden. Administrierende verwalten die Rechte, die über die Rollen der Nutzenden festgelegt sind. Die identifizierten Rollen betreffen dabei verschiedene Aspekte von umwelt.info, wie die Datenbereitstellung, Datennutzung und das CMS.

Administrierende können alle Funktionen des Systems nutzen. Registrierte Nutzende können personalisierte Funktionen in ihrem Profil nutzen. Metadaten-Redakteur\*innen können Metadaten bearbeiten und publizieren. CMS-Redakteur\*innen können die redaktionellen

Inhalte und Einstiegspunkte pflegen. Nicht registrierte bzw. nicht angemeldete Nutzende können im System nach Daten und Inhalten suchen.

Neben der Rollenzuweisung müssen Administrierende in der Lage sein, sich einen Überblick über alle Nutzerprofile zu verschaffen. Ein einzelnes Nutzerprofil muss angezeigt, deaktiviert oder gelöscht werden können.

### **3.3.7 Nutzerprofil**

Das umwelt.info-Portal lässt sich für registrierte Nutzende personalisieren. Registrierte Nutzende haben die Möglichkeit, Benachrichtigungen für gespeicherte Suchen oder geänderte Daten (via E-Mail und im Portal) zu erhalten, Favoriten (Listen) für interessante Inhalte zu erstellen sowie Suchen abzuspeichern. Listen können darüber hinaus auch für andere registrierte Nutzende freigegeben werden. Bei der Registrierung sind folgende Informationen erforderlich: E-Mail-Adresse, Passwort, Vor- und Nachname. Nutzende, die zusätzlich Metadaten bereitstellen wollen, müssen außerdem die Kontaktinformationen ihrer jeweiligen Institution angeben, um eine Verifizierung durch umwelt.info zu ermöglichen. Bereits registrierte Nutzende können sich über eine Schaltfläche im umwelt.info-Portal mit ihrem Nutzernamen und Passwort anmelden. Die Authentifizierungskomponente von umwelt.info unterstützt Single Sign-On, sodass die Anmeldung nur einmal erfolgt und für alle Dienste von umwelt.info gültig ist.

Angemeldeten Nutzenden stehen im umwelt.info-Portal weitere Funktionen zur Verfügung:

- ▶ Verwaltung der personalisierten Inhalte (Favoriten, gespeicherte Suchen) und Anzeige neuer Benachrichtigungen
- ▶ Änderung der persönlichen Profildaten
- ▶ Löschen des eigenen Nutzerprofils.

Die personalisierten Profilfunktionen werden der Benutzeroberfläche durch eine Profilkomponente bereitgestellt, die mit der Authentifizierungskomponente die erforderlichen Berechtigungen abfragt (Server-seitig und Client-seitig). Die via Benutzeroberfläche verfügbaren Funktionen werden entsprechend der zugewiesenen Rolle angepasst.

Benachrichtigungen über Aktualisierungen von gespeicherten Favoriten (z. B. Datensätzen) und Suchen erhalten angemeldete Nutzende im Portal und per E-Mail. Dazu gleicht die Nutzerprofilkomponente die Zeitstempel der letzten Betrachtung (bei jedem Betrachten wird der Zeitstempel aktiviert.) der gemerkten Datensätze und Suchen mit den Zeitstempeln der Metadaten und Anzahl der Ergebnisse ab, um die Anzahl der Aktualisierungen im Portal zu visualisieren und um eine Liste der aktualisierten Datensätze und Suchen bereitzustellen. Eine weitere Funktion prüft periodisch auf Aktualisierungen und erstellt eine Zusammenfassung, die Nutzende per E-Mail erhalten. Dabei können Nutzende Einstellungen über die Zusammenfassungen tätigen (wählbarer Zeitraum, z. B. wöchentlich).

Angemeldete Nutzende können sich am umwelt.info Portal abmelden. Auch hierbei muss die Authentifizierungskomponente Single Logout unterstützen und die Abmeldung an allen beteiligten Komponenten durchgeführt werden. Eine Komponente, die den Anforderungen der Nutzungsverwaltung und des Nutzungsprofils entspricht, wird mit den Komponenten „CMS“ und „Identity Management“ (vgl. 3.4) eingeführt.



### 3.3.8 Programmierschnittstellen

umwelt.info stellt Programmierschnittstellen (Application Programming Interfaces, API) für die maschinelle Nutzung ausgewählter Funktionen des Portals bereit. Dabei sollen Standards helfen, den Arbeitsaufwand für Nutzende so weit wie möglich zu minimieren. Zudem sollen alle Schnittstellen dokumentiert sein. Die API-Komponente in umwelt.info ist auch als einheitlicher Zugangspunkt für Anfragen an das System zu verstehen.

Welche APIs bereitgestellt werden, wurde nicht abschließend definiert und ist im weiteren Projektverlauf zu entscheiden. Beispiele sind:

- ▶ Interoperabilitätsinterfaces: Auf der Basis des Metadata-Index sollten eine oder mehrere der bekannten Schnittstellen für die Interoperabilität mit anderen Metadatenportalen oder Suchmaschinen implementiert werden. Hierzu gehört die "Infrastructure for Spatial Information in the European Community" (INSPIRE) und die OGC API Records-Schnittstelle für die Suche und das Harvesting der Metadaten.
- ▶ Eine serverseitige Schnittstelle für die direkte Suche wird vom Index/Such-Backend bereitgestellt (verbreitet sind zum Beispiel Apache Lucene mit Solr oder Elasticsearch). Durch diesen Ansatz ist der benötigte Funktionsumfang und die Dokumentation der API bereits durch die verwendete Standardsoftware gegeben.
- ▶ Komponenten, wie die Data-Check-In-Komponente, stellen ihre Funktionen auch als API bereit. Zugriffe auf die APIs der einzelnen Komponenten werden über ein API-Gateway angesteuert und müssen genauso wie Zugriffe über das CMS berechtigt sein. Eine Prüfung der Zugriffsberechtigung erfolgt über die Identity Manager-Komponente.

### 3.3.9 Qualität von Metadaten

Im umwelt.info-Portal werden aus verschiedenen Datenquellen (extern/intern) Metadaten mittels Harvester oder Crawler in das System integriert. Diese Metadaten verfügen über unterschiedliche Qualität. Die Metadatenqualität misst sich an einer Reihe von Kriterien für die einzelnen Felder der Metadatenansätze. Die (Meta-)Datenqualität beinhaltet die Aktualität von Datensätzen, Eindeutigkeit, Form der Beschreibungstexte, Kontaktpersonen etc.; generelle Qualitätskriterien betreffen die Vollständigkeit (Befüllung von Feldern).

Die Vielfalt der im Index enthaltenen Metadaten misst die Verteilung oder Anzahl von unterschiedlichen Werten bestimmter Felder. Die Vielfalt der Metadaten beinhaltet vor allem die Felder Inhalt, Datenbereitstellende und Formate. Dabei sind Felder, die für die Suche verwendet werden, von besonderer Bedeutung (z. B. Werte für Facetten).

Zusätzlich zur Metadatenqualität gibt es einige Felder, die indirekt Aussagen über die Qualität der beschriebenen Daten zulassen. Dazu gehören insbesondere „Maschinenlesbar“ oder „Amtliche Datenbereitsteller“<sup>5</sup>. Dabei ist es allerdings nicht das primäre Ziel, die Qualität der Daten zu erfassen, sondern vielmehr die Qualität der Suche sicherzustellen.

Datenqualität und Datenvielfalt sind qualitative Eigenschaften der Metadaten, die mit geeigneter Methodik überprüft werden sollen. Daher wird im umwelt.info-Portal eine Metadata Quality Assessment (MQA)-Komponente eingesetzt, mit der die gesammelten Metadaten aus den verschiedenen Datenquellen geprüft werden.

Exkurs: Zusätzliches Wissen über die jeweiligen Daten kann nur in sehr geringem Umfang zur Sicherung der Qualität der Metadaten in umwelt.info genutzt werden. Datenbereitstellende

---

<sup>5</sup> „Amtliche Datenbereitsteller“ sind Behörden, die Daten im Zuge ihrer Amtsführung erzeugen



haben besonderes Wissen über die Daten und wären dazu in der Lage Informationen, die nicht direkt im Portal einsehbar sind, zu teilen. So könnten Datenbereitstellende möglicherweise Wissen darüber weitergeben, wie die Daten erhoben wurden oder wie diese valide genutzt werden können. Eine Einbindung von Spezialwissen der datenhaltenden Stellen kann nur sehr begrenzt zur Sicherung der Qualität der Metadaten erfolgen und umwelt.info kann keine Fragen zur Erhebung und Nutzung der Daten beantworten. Datenbereitstellenden ist es aber möglich die vorhandenen Felder der jeweiligen Metadatenstandards zu nutzen, um Informationen bestmöglich zu teilen. Die Felder der Metadaten schemata erlauben es zusätzliche Informationen zu den Daten anzugeben oder Hilfsmaterialien zu verlinken, die den Nutzenden die Interpretation und Weiterverwendung der Daten erleichtern können. Daher sollten die datenhaltenden Stellen bei der Veröffentlichung von Metadaten unterstützt werden, um diese Möglichkeiten auszuschöpfen. Zum Beispiel können die Mitarbeiter von umwelt.info die Datenbereitstellenden dabei unterstützen, die passenden Felder im jeweiligen Metadaten schema zu identifizieren.

### 3.4 Komponenten

In der Bausteinsicht erfolgt eine Zerlegung des umwelt.info-Portals in Komponenten und es werden die Beziehungen der Komponenten untereinander dargestellt. So entsteht ein Grundrissplan für die Architektur von umwelt.info. Das System besteht auf oberster Ebene aus den folgenden Komponenten (überarbeitet nach [1]):

**Content Management System (CMS):** Der Einstiegspunkt des Portals ist das Content Management System, das eine Benutzerschnittstelle bereitstellt, über die Nutzende nach unterschiedlichen Daten zu verschiedenen Themenbereichen suchen können. Das CMS speichert die Inhalte der Seiten des umwelt.info Portals und stellt diese den Nutzenden bereit. Dabei sind einige Funktionen (z. B. die Suche) für alle Nutzenden verfügbar und andere nur für angemeldete bzw. berechnigte Nutzende (z. B. Personalisierte Listen bzw. Datenbereitstellen).

**Metadata Harvesting und Crawling:** Das Metadata Harvesting und Crawling ermittelt aus den verschiedenen Datenquellen außerhalb des Systems die Metadaten, indiziert diese und macht sie somit auffindbar.

**Data Check-In:** Mittels der Komponente Data-Check-In können Metadaten für weitere Datenquellen manuell erfasst werden.

**Metadata-Index:** Der Metadata-Index enthält alle Metadaten, die über das Metadata Harvesting integriert oder per Data Check-In eingepflegt wurden (in einer für die Suche optimierten Form). Der Metadaten-Index stellt einen effizienten Indizierungs- und Suchmechanismus bereit und ist in der Lage, das verwendete Metadatenmodell zu interpretieren, bestimmte Suchparameter herauszuziehen und zu indizieren. Der Index kann eine Volltext-Indizierung durchführen.

**Application:** Über Application können die Nutzenden die Daten im umwelt.info Portal visualisieren und daraus neue Erkenntnisse ableiten. Dabei werden den Nutzenden unterschiedliche Anwendungen zur Visualisierung der Daten bereitgestellt. Für die Verwendung der Application müssen sich die Nutzenden am umwelt.info Portal registrieren. Nach der Registrierung stehen den Nutzenden die unterschiedlichen Anwendungen zur Verfügung. Die dazugehörigen Anforderungen sind noch nicht abschließend definiert worden und müssen vor einer Umsetzung im umwelt.info Portal noch näher spezifiziert werden

**API:** Die API stellt Schnittstellen für die maschinelle Nutzung der Funktionen des Portals bereit. Diese sind die Interoperabilitätsinterfaces, eine Identifizierungs- und Permalink-Schnittstelle für

Datensätze, Suchschnittstellen und weitere freigegebene Back-End-Dienste, die von der Benutzeroberfläche oder Nutzenden benötigt werden.

**Identity Management:** Das Identity Management stellt alle notwendigen Funktionen bereit, damit Nutzenden den passenden Zugriff auf die Komponenten des umwelt.info Portals erhalten. Dazu zählen Identifikation und Authentifizierung von Nutzenden, Autorisierung des Zugriffs auf geschützte Ressourcen, und Verwaltung der Nutzenden durch die Portaladministrator\*innen.

**MQA (Metadata Quality Assessment):** Für die Erhebung von Statistiken zur Qualität der im Metadata-Index gespeicherten Metadatenätze wird eine Metadata Quality Assessment (MQA)-Komponente eingesetzt, welche die durch Harvesting und Crawling gesammelten Metadaten prüft. Zusätzlich enthält diese Komponente einen Validator, der von Datenbereitsteller\*innen genutzt werden kann, um Fehler in den eigenen Metadaten zu vermeiden.

**RS (Recommender System):** Auf den Detailseiten zu den Metadatenätzen werden weitere Metadatenätze als Empfehlungen für Nutzende verlinkt, die auf dem jeweils aktuellen Kontext basieren. Dafür wird ein sogenanntes Recommender System (RS) benötigt, das den Kontext als Eingabe entgegennimmt und als Ausgabe Empfehlungen gibt.

**Nutzertracking:** Für die Erhebung von (anonymisierten) Statistiken zur Nutzung im umwelt.info Portal wird eine Nutzertracking-Komponente eingesetzt. Dadurch sollen wesentliche Informationen gesammelt werden, die helfen, das umwelt.info-System für die Nutzenden stetig zu verbessern und gezielte Anpassungen zu entwickeln.

**Data-Source:** Unter Data-Source werden hier die Quellen aller Daten, die durch die Harvester und Crawler oder durch manuelle Eingabe erfasst werden, verstanden. Auf die Quellen der Data-Source wird im Metadata-Index des umwelt.info Portals verwiesen.

Nähere Beschreibungen zu den Komponenten inklusive der Schnittstellen sowie eine weitere Zerlegung in Subkomponenten finden sich in Anhang A.

## 4 Anwendungsmöglichkeiten von Linked Data und KI-Methoden

Der Begriff Künstliche Intelligenz (KI) ist von einem Forschungsgebiet aktuell zu einem überbeanspruchten Trendbegriff geworden, der sehr unterschiedlich interpretiert wird. Beispielhaft seien drei Definitionen genannt. Der Bitkom beschreibt Künstliche Intelligenz als „die Eigenschaft eines IT-Systems, „menschenähnliche“, intelligente Verhaltensweisen zu zeigen“ [8]. Microsoft beschreibt KI als „Technologien, die menschliche Fähigkeiten im Sehen, Hören, Analysieren, Entscheiden und Handeln ergänzen und stärken“ [9]. Das Europäische Parlament definiert KI als „die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren“ [10].

Alle diese Definitionen haben eine Gemeinsamkeit: sie sind sehr offen bzw. uneindeutig formuliert und geben daher nur geringen Aufschluss darüber, was sich genau hinter KI verbirgt. Dieser Bericht hat dabei nicht den Anspruch, eine eigene oder gar bessere Definition zu liefern. Es soll hier nur auf die Problematik über das Verständnis zum Begriff Künstliche Intelligenz aufmerksam gemacht werden, um auch aufzuzeigen, weshalb eventuell an dieser Stelle Themen mit abgehandelt werden, welche nicht von allen Seiten als klassische KI-Elemente betrachtet werden. Als Schnittmenge der KI-Anwendungen werden in diesem Bericht Algorithmen begriffen, welche auf Basis zugrundeliegender Daten Vorhersagen für zukünftige Ereignisse (Texteingaben, Nutzerbedarfe, usw.) treffen und präsentieren bzw. durchführen können [11] [12].

Primär sind die für das umwelt.info geplanten KI-Anwendungen dem Bereich Natural Language Processing (NLP) zuzuordnen. Dieser Bereich befasst sich damit, wie Computer algorithmisch natürliche Sprache verarbeiten können [13] [14]. Hierunter fallen unterschiedlichste Elemente, vom Verarbeiten kurzer geschriebener Sätze bis hin zur Eingabe mittels gesprochener Worte. Gemeinsamkeit aller NLP-Ansätze ist der Zweck Intentionen, Hauptaussagen und Kontext aus natürlichsprachlichen Elementen herauszufiltern und Vorhersagen bzw. Entscheidungen daraus zu entwickeln. Dies betrifft insbesondere die hier vorgestellten Themen wie Sucheingabevorschläge, Empfehlungsalgorithmen, Question-Answering Systeme, sowie Teile der KI-Suche und der Metadaten-Qualitätssicherung. Etwas besonders gestalten sich dabei die Empfehlungsalgorithmen, da diese häufig zusätzlich auf unterschiedlichen Aspekten, wie z. B. Nutzerprofilen, betrachteten Datensätzen usw. aufbauen [15].

Um abzusichern, dass die KI-Anwendungen möglichst optimal und einfach nutzbar gestaltet werden und nur Anwendungen in die Konzeptionierung des Portals aufgenommen werden, die einen eindeutigen Mehrwert versprechen, wurden zwei acht-stündige KI-Workshops durchgeführt. In den Workshops wurden zunächst die KI-Methoden kurz vorgestellt und deren mögliche Einsatzbereiche skizziert. Anschließend wurden die Einsatzszenarien diskutiert und die Einsatzfelder eingegrenzt. Dazu wurden in den Workshops User Journey Maps (UJM) erstellt, Lösungsansätze der festgestellten Aufgaben und Anforderungen durch ein Round-Robin Brainstorming erarbeitet und diese in einer Entscheidungsmatrix bewertet. Die Ergebnisse dienten als Basis für die technische Konzeption, die in der Systemarchitektur dokumentiert ist.

Methoden aus dem Bereich der Künstlichen Intelligenz und die Nutzbarmachung von Linked Data wurden untersucht, um die Nutzbarkeit des Portals zu verbessern. Ziel ist es, die Suchergebnisse zu den Sucheingaben der Nutzenden zu optimieren und intelligente Hilfestellung zu geben. Durch den Einsatz dieser Techniken ergeben sich darüber hinaus neue Ansätze für die Verbesserung der Datenqualität. Für die Umsetzung von Funktionen im Zusammenhang mit „Spatial Data on the Web“ (SDW; z. B.: [16] und „Linked Data“ (LD) [17]),

sollte die Architektur konsistent sein mit den Prinzipien des Web, aber auch mit den Prinzipien von Linked Open Data [18].

Zu den nachfolgend dargestellten Anwendungsfeldern werden entsprechende Komponenten eingeführt, deren Beschreibungen im Anhang A zu finden sind. Die Funktionen 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10 sind dabei im oben umrissenen Sinne dem Bereich KI zuzuordnen, wobei die Funktionen 4.3 (Lösungsansatz B), 4.6, und 4.9 insbesondere auf Methoden aus dem Bereich Linked Data setzen.

## 4.1 Sucheingabevorschläge während der Eingabe

Eine erste Unterstützungsmöglichkeit mittels KI ist während der Eingabe in den Suchschlitz möglich. Während der Eingabe einer Suche bekommen Nutzende Vorschläge angezeigt, die zur bisher eingegebenen Suche passen und die über den Cursor ausgewählt werden können. Dabei soll die Autovervollständigung sinnvolle Suchen vorschlagen. Vorschläge sollen zur bisherigen Eingabe passen, relevant für die Nutzenden sein und nicht zu trefferlosen Ergebnisseiten führen. Es lassen sich verschiedene Formen von Suchvorschlägen unterscheiden. Eine Autokorrektur kann Fehler (z. B. Tippfehler) beheben oder es können Vorschläge für eine Vervollständigung des aktuellen Wortes oder für eine Vervollständigung des aktuellen und des nächsten Wortes gegeben werden. Es können aber auch vollständige Sucheingaben als Vorschläge gegeben werden, wobei Formulierungen die bisher eingegebenen Wörter und Zeichen enthalten können, aber nicht müssen (z. B. könnte für die unvollständige Suche „Fluss“, der Vorschlag „Binnenschiffahrt in Deutschland“ generiert werden). Dabei schließen sich diese Möglichkeiten nicht gegenseitig aus und sind bereits Stand der Technik. Eine Vervollständigung der Sucheingabe mit einer genauen „Vorhersage“ kann auch als solche auch differenziert werden<sup>6</sup>.

Aufgrund der Verwendung „während der Eingabe“ müssen Vorschläge sehr performant und häufig (im Idealfall nach jedem Tastenanschlag) von der verantwortlichen Komponente geliefert werden können. Suchvorhersagen können inhaltsbasiert (indexierte Textdaten) oder nutzungsbasiert (getrackte Sucheingaben als Daten) sein, auch hybride Ansätze sind möglich. Vorhersagen können allein auf einem Index basieren, oder durch den zusätzlichen Einsatz von Natural Language Processing (NLP) generiert werden.

**Lösungsansatz:** Es wird eine „Search Prediction“-Komponente eingeführt, die von der Benutzeroberfläche während der Eingabe angefragt werden kann. Um die benötigte Performanz zu erreichen können Vorschläge vorbereitet und selbst in einem Index vorgehalten werden.

Neben einer inhaltsbasierten Basisumsetzung, die etwa die durch spezielle Indexabfragen die „forward“ und „reverse Index“-Funktionalität der Suchmaschine kombiniert, sind weitere Ansätze denkbar. Mögliche NLP-Methoden basieren zum Beispiel auf Logik, Algorithmen, Knowledge Graphen oder Sprachmodellen (siehe Anhang A).

## 4.2 Ähnliche Suchen

Vorschläge zu verwandten bzw. ähnlichen Suchen werden niederschwellig immer nach einer Suche präsentiert, zeigen sich jedoch besonders präsent, wenn die Suche keine Treffer geliefert

---

<sup>6</sup> Google beschreibt diese Art von Autovervollständigung als „Autocomplete Predictions“, um zu verdeutlichen, dass es sich um Vorschläge handelt, die nicht nur helfen die Suche einzugeben (vervollständigen), sondern auch, um andere, möglicherweise treffendere, Suchen zu formulieren.

<https://blog.google/products/search/how-google-autocomplete-works-search/>

hat. Für eine gegebene Suche sollen den Nutzenden weitere Suchvorschläge gemacht werden. Diese Funktion ähnelt der Vorhersage von Sucheingaben. Die Aufgabe, die es dabei zu lösen gilt, unterscheidet sich aber in einigen Detailpunkten. Insbesondere hinsichtlich Laufzeitanforderungen. Auch bei Eingabe und Ausgabe sind Unterschiede zu beachten. Insbesondere muss keine Vervollständigung oder Vorhersage der Sucheingabe erreicht werden.

Vorschläge können auf zwei unterschiedlichen Wegen produziert werden: entweder basierend auf vergleichbaren Suchen anderer Nutzenden oder basierend auf einer NLP-Analyse der Sucheingabe mit Bezug auf die Inhalte im Metadatenindex. Ersteres würde voraussetzen, dass ein gewisser Grad an Nutzenden-Tracking stattfindet. Dies kann entweder durch Cookies erfolgen oder durch Zustimmung bei Anmeldung im Portal. Die Verwendung der Suchhistorien von angemeldeten Nutzern setzt dabei aktivere Nutzende voraus, die zudem Ihre Zustimmung gegeben haben. Allerdings wurde in den KI-Workshops die Verwendung von Cookies allgemein kritisiert (störende Elemente beim Aufrufen der Seite, eventuelle Unsicherheit bei rechtlichen Änderungen). Alternativ oder parallel bietet sich daher der Weg über die NLP-Analyse an, da dabei eine reine Analyse der vorliegenden Eingaben – vergleichbar mit der Autovervollständigung oder dem Thema „Suche Erweitern“ – erfolgen muss.

**Lösungsansatz:** In der Bausteinsicht wird eine „Search RS“-Komponente eingeführt, die von der Benutzeroberfläche neben einer Suchanfrage genutzt werden kann. In einer Basisumsetzung wird die Funktionalität des Index genutzt. Darüber hinaus werden für diese Komponente verschiedene weitere Ansätze und Optionen beschrieben (ähnlich Lösungsansatz 4.1; siehe Anhang A).

### 4.3 Suche Erweitern

Es ist davon auszugehen, dass bei bestimmten Sucheingaben entweder zu wenige Treffer gefunden werden oder sich zwar viele Treffer ergeben, diese aber nur eine geringe Relevanz zur Suchanfrage haben. Ersteres kann daran liegen, dass die Sucheingabe zu speziell, bzw. mit zu vielen einschränkenden Begriffen formuliert wurde, letzteres daran, dass sehr allgemein bzw. mit zu wenigen einschränkenden Begriffen gesucht wurde. In beiden Fällen können KI-Algorithmen unterstützen, indem alternative Begriffe vorgeschlagen werden. Eine „Sucherweiterung“ ist ein Anwendungsgebiet für Semantische Analysen und Linked Data. Grundsätzlich ist hier gemeint, dass die für die Suche verwendeten Suchworte (allgemeiner: Konzepte) durch speziellere (oder allgemeinere) Konzepte ersetzt bzw. erweitert werden. Eine „Sucherweiterung“ kann auf mehrere Arten realisiert werden (Lösungsansatz A und B im Folgenden sowie Anhang A).

**Lösungsansatz A (Logische Operationen in der Suchanfrage):** Durch Bezugnahme auf die Eingaben der Nutzenden im Suchschlitz, können passende, verwandte Begriffe oder Konzepte ermittelt und vorgeschlagen werden. Dieses löst eine „UND/ODER-Suche“ (ähnlich einer Google-Suche) aus, sofern die Suchworte nicht in Hochkommata stehen (s. u.). Dabei werden die Treffer, die alle Suchworte möglichst gut abdecken, möglichst weit oben angeordnet, während Treffer die nur einzelne Suchworte abdecken, sich weiter unten in der Trefferliste wiederfinden. Suchworte, die in Hochkommata stehen, werden mit „und“ verknüpft, d. h. die Treffer, die angezeigt werden, enthalten mindestens die in Hochkommata stehenden Suchworte. Die Begriffe, die für eine Sucherweiterung benötigt werden, werden durch spezielle Indexabfragen (ähnlich Lösungsansatz 4.1) ermittelt, die Zusammenhänge zwischen vorkommenden Begriffen im Index analysieren. Diese aus der aktuellen Eingabe im Suchschlitz abgeleiteten Begriffe werden in der 1. Zeile unter dem Suchschlitz dargestellt und können dann zur Sucherweiterung ausgewählt werden. Dadurch werden sie dem Suchschlitz hinzugefügt (in Hochkommata, also

mit einer UND-Operation konjugiert). In der 2. Zeile unter dem Suchschlitz befinden sich die Worte, die bereits im Suchschlitz verwendet werden und wieder entfernt werden können.

**Lösungsansatz B (Filtern von Konzepten aus Thesauri):** Neben dem Suchschlitz können die Facetten (als Filter) herangezogen werden, um die Suche einzuschränken bzw. zu erweitern. Diese filtern allerdings „hart“ Treffer aus der Treffermenge. Besonders die Facetten, die auf Thesauri (Taxonomien) basieren, eignen sich für eine (schrittweise) Spezialisierung der Suchbegriffe (Konzepte). Ein Thesaurus gibt dabei die hierarchische Anordnung von Konzepten zu einem bestimmten Thema vor. So gibt es etwa in der UMTHEs Taxonomie das Konzept der „Umweltverschmutzung“ das als spezialisierte Konzepte u.a. „Ölverschmutzung“ und „Luftverschmutzung“ enthält. Die Verwendung des Begriffes „Ölverschmutzung“ anstelle des allgemeineren Begriffes der „Umweltverschmutzung“ würde dann zu einer Einschränkung der bisherigen Treffermenge führen, da im allgemeineren Fall von „Umweltverschmutzung“ neben diesem Begriff zusätzlich nach „Ölverschmutzung“ und „Luftverschmutzung“ gesucht würde. Eine Spezialisierung bei der Suche macht aber nur Sinn, wenn danach mindestens noch 1 Suchtreffer vorliegen würde. Analog zur Verwendung des Thesaurus zur Spezialisierung kann dieser auch für die Verallgemeinerung der Suche verwendet werden: wenn zuvor ein oder mehrere spezialisierte Konzepte verwendet wurden, kann dafür ein allgemeineres Konzept angewählt werden. Für eine Auswahl spezialisierter (oder allgemeinerer) Konzepte müsste die Baumstruktur des Thesaurus traversierbar sein, damit der Anwendende das (oder die) Konzept(e) finden kann, das (die) für die Suche selektiert werden soll. Hierzu müsste die Benutzeroberfläche je eine Facette pro verwendeten Thesaurus einblenden. Dieses Oberflächenelement könnte eine spezielle Navigationsfunktion bereitstellen: wenn das Kindelement des aktuellen Wurzelknotens ausgewählt wird, wird das Element in der Visualisierung zum Wurzelknoten und dessen mögliche "Kinder" werden angezeigt und können ausgewählt werden.

Grundlage für die Thesaurus-basierte Suche ist, dass die Metadaten mit einem oder mehreren Begriffen aus der Taxonomie verschlagwortet sind, praktisch also Links zwischen ihnen und Konzepten aus der Taxonomie erstellt wurden. Dieses ist entweder bereits in den originären Metadaten gegeben oder die Harvester bzw. Crawler müssten hierfür sorgen. Dazu können verschiedene Methoden verwendet werden. Das können auch KI-Methoden sein, die auf der Basis von Trainingsdatensätzen gelernt haben, Kurzbeschreibungen (wie in den Metadaten vorliegend) bestimmten Konzepten einer Taxonomie zuzuweisen. Letzteres wird bereits im Umweltbundesamt (UBA) für Forschungsberichte im Zusammenhang mit der UFORDAT [19] praktiziert. Es könnte auch versucht werden, etwa geographische Koordinaten mit (einem) Begriff(en) aus einer geographischen Taxonomie zu verknüpfen und diese(n) Begriff(e) für die Annotation zu verwenden (inverses Geocoding).

#### 4.4 Ähnliche Datensätze

Die Empfehlung für ähnliche Datensätze erfolgt für Nutzende nach „Auffinden“ der gesuchten Daten. Die Intention dabei ist es, Nutzenden möglichst umfassend Informationen zum gesuchten Thema zur Verfügung zu stellen. Hierbei können neben einer Ähnlichkeit des gefunden Datensatzes und anderen im Portal auffindbaren Datensätzen, zusätzlich noch Informationen aus der Sucheingabe verwendet werden. Ebenfalls können hierfür noch Informationen aus den Interessen des Nutzenden hinzugezogen werden. Dazu könnten Informationen aus dem Nutzerprofil, sowie vergangene Suchen oder zuvor betrachtete Datensätze genutzt werden. Im Ergebnis werden auf den Detailseiten zu den Metadatenätzen weitere Metadatenätze als Empfehlungen für Nutzende verlinkt, die auf dem jeweils aktuellen Kontext basieren.



**Lösungsansatz:** Für die Ermittlung passender Datensätze wird ein sogenanntes Recommender System (RS) als Komponente benötigt, die den „Kontext“ (z. B. letzte Suche, aktuell angezeigte Datensätze) als Eingabe entgegennimmt und als Ausgabe Empfehlungen gibt, welche Metadatenätze zum gegebenen Kontext außerdem passen. Eine einfache Lösung basiert auf Indexabfragen und wird in Anhang A (Recommender System – Basisimplementierung, Inhalts-/Indexbasiert) beschrieben. Der im Designkonzept dargestellte, alternative Ansatz berücksichtigt auch die Metadatenätze der anderen Nutzenden. Dieser Ansatz wird im Anhang A unter „Recommender System – Merklisten und Collaborative Filtering“ weiterverfolgt.

## 4.5 Antworten auf konkrete Fragen

Für Eingaben im Suchschlitz, die als Frage formuliert sind, soll eine Funktion „Question-Answering“ eingesetzt werden, die für eine abgesendete Suche, eine konkrete Antwort bereitstellt. Antworten werden in der Nutzeroberfläche zusätzlich zur Suchergebnisliste angezeigt und sollen diese lediglich ergänzen. Nutzende werden bei den meisten Suchanfragen mit den wichtigsten Informationen in den ersten Zeilen bedient, teilweise auch mit Bildern, Fotos oder Graphiken. Um diese Informationen zu liefern, analysiert eine entsprechende Komponente hinterlegte Daten, um eine möglichst gute Übereinstimmung mit der Suchanfrage zu finden [20] [21] [22].

**Lösungsansatz:** Antworten können mittels zwei verschiedener Techniken ermittelt oder generiert werden: Mit dem „Passage Retrieval“ wird in einem indexierten, geeigneten Textkorpus nach Antworten gesucht. Hierbei werden die gefundenen Textfragmente als Antwort ausgegeben. Ein alternativer Ansatz leitet Antworten aus einem geeigneten Knowledge Graphen ab. Für die hier beschriebene Funktion des „Question Answering“ wird eine passende Komponente „Q/A-Retrieval“ entwickelt, die auf NLP-Methoden und ein entsprechendes Framework zurückgreift. Für die Funktionalität dieser Komponente ist es sinnvoll, Suchanfragen zunächst zu klassifizieren, um den „Fragetyp“ zu ermitteln. Darüber hinaus sollte auch anhand der Treffermenge für eine Suche oder Frage bewertet werden, ob die ermittelte Antwort sinnvoll ist. Weitere Details werden in Anhang A dargestellt.

## 4.6 Erweiterte Antworten / Mashups

Bei erweiterten Antworten in Form von Mashups handelt es sich um Web-Seiten, die verschiedene Informationen kombinieren, um einen schnellen Überblick über ein Thema zu bieten. Dadurch wird es nicht ausschließlich den Suchenden überlassen die verlinkten Informationen sukzessive aufzusuchen, um sich daraus die wichtigsten Informationen zu extrahieren.

**Lösungsansatz:** Grundlage hierfür ist der aufgebaute Knowledge Graph mit verlinkten Ressourcen. Konkret kann hierfür vom System beispielsweise nach DCAT-AP Ressourcen (DCAT-AP Graphen) im Metadaten-Index gesucht werden, die möglichst gut auf die Suchanfrage passen (z. B. hohe Ähnlichkeit bei Begrifflichkeiten, Raumbezug, Zeit, ...). Im Ranking kann auch berücksichtigt werden, dass Ressourcen, die viele eingehende Links anderer Ressourcen aufweisen (was darauf hinweist, dass die gefundene wohl einen großen Beitrag zum angefragten Sachverhalt aufweist) möglichst weit nach oben wandern. Dann kann versucht werden, den Graphen der durch die Attribute des am höchsten gerankten Treffers sowie der mit diesem Treffer verlinkten Ressourcen aufgespannt wird, textlich in einer Web-Seite aufzubereiten, die neben textlichen Informationen auch verschiedene andere Informationen kombiniert (sofern sinnvoll und vorhanden z. B. die Darstellung der räumlichen Ausprägung in einer Karte, Bilder, Link zu Videos etc.).

## 4.7 Intelligente Suche

Das Ziel einer Suchmaschine ist es, den Nutzenden zu helfen die für ihren Informationsbedarf relevanten Informationen zu finden. Eine intelligente Suche ermöglicht dabei die Sortierung von Treffern nach Relevanz (vgl. Designkonzept, Sortierung nach Relevanz) und bietet optimale Unterstützung beim Auffinden der gewünschten Informationen. Im Rahmen einer intelligenten Suche sollen zunächst die in diesem Kapitel beschriebenen Funktionen für die Unterstützung der Nutzenden eingesetzt werden. Das bedeutet, dass KI-Funktionen Hilfestellungen anbieten, ohne die Sucheingabe im Hintergrund, „unsichtbar“ für die Nutzenden, zu verändern. Es sollten darüber hinaus die technischen Möglichkeiten genutzt werden die Relevanz der Suchtreffer hinsichtlich der Sucheingabe zu optimieren. So ist es sinnvoll, die Sucheingaben zu analysieren, um etwa dem Rankingalgorithmus zu erlauben auch nach Konzepten zu suchen, statt nur nach Schlagwörtern. Das Ziel ist es also Suchanfragen zu „verstehen“, um die bestmöglichen Treffer zu liefern. Hierbei darf für die Nutzenden jedoch nicht der Eindruck entstehen, dass Ihre Suchen verändert werden.

**Lösungsvorschlag:** Die Komponente „KI-Suche“ sollte für die Suche im Metadaten-Index genutzt werden (anstelle der direkten Suche auf dem Metadaten-Index). So können nachgelagert (also nach der Sucheingabe), verschiedene NLP-Methoden (etwa unter Nutzung eines NLP-Frameworks) und spezielle Anfragetechniken gegen den Metadaten-Index genutzt werden. Durch den Einsatz von Optimierungsmethoden für das Ranking kann die Relevanz der Suchtreffer im Betrieb erhöht werden. Dabei kann etwa der Einsatz eines „Learning to Rank“-Verfahrens anstelle des Standardrankings geprüft werden. Mit weiteren Informationen, die die KI-Suche bereitstellt (etwa auf Basis des Suchverlaufs des Nutzenden), kann die Benutzeroberfläche auch entscheiden, welche unterstützenden Funktionen für die Nutzenden möglicherweise gerade besonders sinnvoll sind, falls diese nicht jederzeit eingeblendet werden sollen.

## 4.8 Qualitätssicherung von Metadaten

Die Unterstützung bei der Qualitätssicherung von Metadaten mittels Methoden der KI ist ein vielschichtiges und komplexes Thema, da es viele Anknüpfungspunkte gibt, die teils mit deutlichem Forschungsbedarf verbunden sind. Da viele der Funktionen des Portals auf Metadaten aufbauen, ist eine hohe Qualität dieser Metadaten eine Voraussetzung für eine erfolgreiche Umsetzung. Aus diesem Grund wurde dem Thema auch im zweiten KI-Workshop während der Konzeptionierung viel Zeit eingeräumt. Als Ansatzpunkte für KI in der Metadaten-Qualitätssicherung wurden zwei Bereiche herausgearbeitet: (a) Unterstützungen während der Eingabe des Metadaten-Formulars und (b) Überprüfung bestehender Metadaten. Einfache Qualitätsprüfungen der Metadaten wurden in Kap. 3.3.9 beschrieben (etwa die Prüfung auf Befüllung eines Feldes) und werden einer MQA-Komponente realisiert (siehe Anhang A). Dabei handelt es sich um Prüfungen auf „technische“ oder formale Korrektheit. Es soll darüber hinaus versucht werden weitere, zum Beispiel inhaltliche Fehler zu erkennen. Denkbar ist hierbei der Einsatz von NLP und ein Abgleichen der Felder untereinander und ein Abgleichen der Metadaten mit den Daten selbst. Diese Funktion soll insbesondere für die Eingabemaske eingesetzt werden.

**Lösungsansatz:** Die MQA-Komponente kann erweitert werden, um entsprechende Index- und NLP Funktionen einzusetzen. Dazu wird in der Bausteinsicht die KI-MQA-Komponente, anstelle der MQA-Komponente eingeführt. Diese Komponente erhält mehr Möglichkeiten, insbesondere den Zugriff auf die (externen) Daten selbst. Einige Mögliche Ansätze für inhaltliche Prüfungen werden bei der Beschreibung der KI-MQA-Komponente (siehe Anhang A) genannt.



## 4.9 SPARQL-Client

Ein wichtiger Anwendungsbereich für Linked Data sind sicher die flexiblen Möglichkeiten zur Erzeugung komplexer Abfragen auf den RDF-Daten. Diese werden mittels (Geo-)SPARQL durchgeführt. Unter Kenntnis der verwendeten Taxonomien und Ontologien lassen sich mit (Geo-)SPARQL sehr präzise Anfragen stellen.

**Lösungsansatz:** Für die Eingabe einer SPARQL Abfrage und die anschließende Darstellung der Suchergebnisse wird eine Komponente für die Benutzeroberfläche benötigt. Da (Geo)SPARQL Anfragen nur von Anwendenden gestellt werden können, die über gewisse Kenntnisse von Linked Data, RDF, (Geo-)SPARQL sowie den verwendeten Taxonomien und Ontologien besitzen, sollte ein (Geo-)SPARQL-Client nicht zu offensiv in der Benutzeroberfläche in Erscheinung treten, sondern eher einer Gruppe von „Erweiterten Funktionen“ zugeordnet sein. Es sollte sich dabei um eine Art Editor-Fenster handeln, dass mit einer ausführbaren Beispiel-(Geo-)SPARQL-Anfrage vorbelegt ist. Nach dem Editieren der Anfrage, sollte diese ausgeführt und die Ergebnisse dargestellt werden können (ähnlich wie es in anderen Portalen, etwa dem Europäischen Daten Portal möglich ist<sup>7</sup>).

## 4.10 Registrierung in der Linked Open Data Cloud

Ein nicht unmittelbarer Vorteil für Linked Data (der sich über einen längeren Zeitraum aber als solcher herausstellen kann) ist das Integrieren der eigenen Linked Data basierten Metadaten in die Linked Open Data Cloud (LODC)<sup>8</sup>. Die LODC ist ein Netzwerk von mehr als 1300 registrierten verteilten Datensätzen aus verschiedenen Domänen. Die LODC visualisiert Datensätze und deren Links. Man kann seine Daten zur LODC hinzufügen und sie so für semantische Clients sichtbar und abfragbar machen. Hierzu müssen die eigenen Daten den Linked-Data-Prinzipien genügen.

## 4.11 Weitere KI-Methoden

Neben den zuvor beschriebenen Methoden bzw. Anwendungsfällen wurden in der Konzeption auch Methoden aus dem Bereich KI diskutiert, für die ein Einsatz in umwelt.info (vorerst) nicht als sinnvoll erscheint.

Dazu zählen z. B. Chatbots. Als Ergebnis des Workshops wurde ermittelt, dass ein Chatbot im klassischen Sinne für das umwelt.info Portal nicht erwünscht ist. Insbesondere negativ erachtet wurden als Pop-Up erscheinende Dialoge mit künstlich vorgetäushtem menschlichem Erscheinungsbild sowie nachgespielte Dialoge und Dialogsteuerung. Außerdem wurden die zu befürchtenden hohen Wartungsaufwände eines solchen Chatbots als negativ bewertet.

Die Möglichkeit der Einbindung einer Spracheingabe in der Suchmaske bzw. einer Sprachsteuerung des Portals wurde angedacht und im zweiten KI-Workshop diskutiert. Hierbei wurden zwei Ebenen bedacht, zum einen die Einbindung als direkte Funktion auf der Webseite, zum anderen die Anbindung an Smart Speaker bekannter Hersteller. Ersteres wurde unter der Prämisse verworfen, dass eine solche Einbindung größere Probleme mit bereits optimierter Hardware von Personen mit Einschränkungen produzieren könnte. Damit wäre die Einbindung eventuell kontraproduktiv für ihren Hauptzweck. Die Anbindung an Smart Speaker wurde nicht vollständig verworfen, jedoch mit sehr geringer Priorität belegt.

KI-Anwendungen für Customer Intelligence sind allgemein als Teilbereich der Recommender Systeme zu betrachten [23]. In der Diskussion im Projekt, sowie in diesem Bericht sollen solche

---

<sup>7</sup> <https://data.europa.eu/data/sparql>

<sup>8</sup> <https://lod-cloud.net>

Anwendungen als Recommender System verstanden werden, welches sich primär auf einzelne Nutzende als Individuum orientiert. Das bedeutet, dass als primäre – oder sogar einzige – Datenquelle für die Informationen und Empfehlungen aus einer KI-unterstützten Customer Intelligence die Angaben sowie das Verhalten von Nutzenden verwendet werden würden. Der Vorteil ist, dass keine fremden Daten Verwendung finden müssen und damit die „eigenen“ Daten auch nicht für die Auswertung „anderer“ Nutzender hinzugezogen wird. Der Nachteil ist, dass dies bedeutet, dass sehr viele Informationen, sowie Analysen der Verhaltensweise einzelner Nutzender notwendig sind, um das System überhaupt nutzen zu können. Damit wäre diese Implementierung nur für wenige, das System intensiv Nutzende von Interesse.

Für die genannten Anwendungsfelder wurde entschieden, die Möglichkeit der Anbindung hier zu benennen, jedoch keine weitere technische Analyse durchzuführen und dementsprechend auch keine technische Konzeption zu erstellen. Bei Bedarf kann dies in einer späteren Version nachgeholt werden.

## **4.12 Workshopkonzept für begleitende Workshops während der Umsetzung**

Die schnellen Entwicklungen im Bereich der künstlichen Intelligenz (KI) zeigen die Notwendigkeit stetig Lösungen zu bekannten Problemen zu hinterfragen und ggf. zu optimieren. Um trotz allem zukünftige Nutzende nicht zu übergehen und das Projektteam im Einklang mit diesen agieren zu lassen, werden interaktive Workshops während der Umsetzung empfohlen. So können Möglichkeiten neuer Technologien mit den Anforderungen abgeglichen und neue Anforderungen aufgenommen werden. Im Folgenden werden Workshopmethoden für digitale oder Präsenz-Workshops beschrieben, sodass Nutzende erfolgreich in die Entwicklung von umwelt.info miteinbezogen werden können. Das beschriebene Vorgehen wurde im Umsetzungskonzept bereits angewendet.

Ein Workshop gliedert sich in vier Teile. Im Anschluss an ein kurzes Aufwärmen und Vorstellen wird ein kurzes Problembrainstorming durchgeführt. Anschließend werden die KI-Ansätze, die in dem Workshop diskutiert werden sollen, präsentiert, sodass alle Teilnehmenden eine ähnliche Wissensgrundlage haben. Im letzten Schritt werden die Möglichkeiten dieser Ansätze mit den identifizierten Problemen in Verbindung gebracht und diese nach Kosten-Nutzen bewertet.

Teilnehmende des Workshops sollten aus ausgewählten Stakeholdern, wie Nutzenden des Systems aus verschiedenen Nutzungsgruppen, KI-Fachleuten, Entwicklenden und Moderierenden bestehen. Wenn es nicht möglich ist, Nutzende zu dem Workshop einzuladen, sollte auf die in der Machbarkeitsstudie erarbeiteten Personas hingewiesen werden. Teilnehmende können dann während des Workshops die Perspektive von verschiedenen Personas einnehmen und so besser aus Sicht von Nutzenden agieren.

Um die Aufmerksamkeit aller zu erhalten, kann zu Beginn der Energizer Touch Blue durchgeführt werden. Hierbei geht es darum, so schnell wie möglich ein Gegenstand in der gesagten Farbe zu finden bzw. in die Kamera zu halten. Die Person, die als letztes einen Gegenstand gefunden hat, definiert die nächste Farbe. Anschließend wird eine kurze Vorstellungsrunde durchgeführt, in der alle Teilnehmenden sich kurz vorstellen und von ihren Erfahrungen mit KI berichten.

Um Probleme und ihre Auswirkungen zu identifizieren wird die 3W-Methode empfohlen. In drei kurzen Sequenzen werden die Teilnehmenden gebeten, zu einem bestimmten Thema Ihre Gedanken aufzuschreiben. Für jede Sequenz werden fünf Klebezettel verteilt und sieben Minuten Zeit gegeben. Anschließend werden die Teilnehmenden gebeten, die geschriebenen Zettel vorzustellen und an die Wand bzw. auf das virtuelle Board zu ziehen. Passt der Zettel zu

einem vorherig vorgestellten, sollen diese in räumliche Beziehung gesetzt werden. Sollten keine realen Nutzenden an dem Workshop teilnehmen, sollten im Vorhinein zu dieser Übung die Personas unter den Anwesenden aufgeteilt werden. Folglich können so diverse Aspekte identifiziert werden.

In der ersten Sequenz (What?) wird berichtet, welche Probleme beobachtet wurden und was am jetzigen System stört. Hierbei gilt es, die Pain Points so gut wie möglich zu beschreiben. In der zweiten Sequenz (So what?) wird beschrieben, warum diese Beobachtungen und Probleme wichtig sind zu behandeln. Was sind die Folgen dieser Aspekte? Während dieser Sequenz dürfen Teilnehmende auch gerne auf die Zettel anderer eingehen. Während der dritten Sequenz (Now what?) werden verschiedene Handlungsmöglichkeiten für die zuvor erarbeiteten Aspekte erstellt. Ebenso wie in der zweiten Sequenz dürfen Teilnehmende auch auf Zettel anderer eingehen.

Im nächsten Teil des Workshops soll den Teilnehmenden Input in Form von Präsentation von verschiedenen (neuen) KI-Methoden geliefert werden. Nach den Präsentationen werden die Teilnehmenden in Gruppen von drei bis vier Personen eingeteilt. Die Cluster, die in der letzten Sequenz der 3W-Methode gebildet wurden, werden nun so unter den Gruppen aufgeteilt, dass jede Gruppe thematisch ähnliche Cluster bearbeitet. In einer Round-Robin (Brainwriting) Session werden nun Lösungen zu den verschiedenen Clustern entwickelt. Diese Lösungen sollen auf den zuvor gehörten KI-Methoden basieren. Während der Round-Robin Session bekommen alle Teilnehmenden aus einer Kleingruppe einen (virtuellen) Zettel, auf dem eine Idee innerhalb von zwei Minuten generiert werden soll. Danach werden die Zettel an die nächste Person weitergegeben. Diese kann dann entweder die Idee der vorigen Person erweitern oder eine neue Idee aufschreiben, die thematisch zu dem Cluster passt. Nach mindesten vier Runden werden in jeder Gruppe die besten drei Ideen gewählt und anschließend im Plenum vorgestellt.

Im letzten Teil des Workshops werden in einer Kosten-Nutzen Matrix die Ideen der Gruppen bewertet. Hierbei bildet die eine Achse der Matrix die Bedeutung für umwelt.info ab die andere die technische Machbarkeit bzw. den Aufwand. In diesem Teil sollten die Personen, die die Präsentation gehalten haben zur Bewertung aktiv beisteuern und ggf. Rückfragen stellen beziehungsweise Rückfragen der Teilnehmenden beantworten.

## 5 Agiler Prozess für umwelt.info

Für die Realisierung von umwelt.info wurde ein agiles Vorgehensmodell entworfen. Die Ergebnisse der Machbarkeitsstudie [1] zeigen, dass dieses Projekt eine außerordentlich hohe Komplexität aufweist und dass mehrere Organisationen mit der Entwicklung betraut und vom Betrieb betroffen sind. Des Weiteren ist zu erwarten, dass das Anforderungsprofil noch geändert und erweitert wird.

All dies lässt ein agiles Vorgehensmodell besonders geeignet für die Entwicklung des Vorhabens erscheinen. Für den Entwurf wurden mit Scrum, Kanban und Extreme Programming verschiedene gängige Modelle verglichen und bewertet. Die Beschreibung dieser Vorgehensmodelle findet sich im Anhang B, eine Zusammenfassung im folgenden Kap. 5.1.

Als Ergebnis des Vergleichs der verschiedenen Vorgehensmodelle ergibt sich, dass Scrum am besten als Grundlage für die Realisierung von umwelt.info geeignet ist.

Es ist nicht zu erwarten, dass ein bestehendes Vorgehensmodell unverändert für die Realisierung eines Projekts der Größenordnung von umwelt.info verwendet werden kann. Dies trifft auch auf Scrum zu, sodass die Regeln und Inhalte dieses Modells entsprechend den Anforderungen von umwelt.info angepasst werden müssen. Diese Anpassung wird ausführlich im Anhang B und zusammenfassend im Kap. 5.2 behandelt.

Ein agiles Vorgehen bei der Projektumsetzung erfordert die Schaffung geeigneter vertraglicher Grundlagen. Aus diesem Grund wurde eine Rahmenvereinbarung entworfen, die es ermöglicht, das beschriebene Vorgehensmodell im Auftrag des Umweltbundesamtes anzuwenden. Der Entwurf der Rahmenvereinbarung liegt beim UBA vor und kann dort angefragt werden. Die Beschreibung des möglichen Zusammenwirkens dieser Vereinbarung mit dem Vorgehensmodell ist in Anhang B beschrieben.

### 5.1 Agile Vorgehensmodelle

Der wesentliche Grund für die Bevorzugung eines agilen Vorgehensmodells gegenüber einem traditionellen liegt in der Komplexität des Vorhabens. Komplexität meint hier, dass die zu lösenden Probleme in einem System eingebettet sind, welches durch zahlreiche Akteure, Abhängigkeiten und Einflussfaktoren keine reine Analyse und somit Prognose der Schritte zur Lösung zulässt [24]. Im Kontext von Softwareentwicklung entsteht Komplexität zum Beispiel durch die Dynamik von Anforderungen, sich ändernde technische sowie rechtliche Rahmenbedingungen oder durch das Zusammenwirken und die Interaktion von Menschen, Teams und Organisationseinheiten [25, pp. 33-50].

Agile Vorgehensmodelle adressieren die Herausforderungen eines komplexen Problemraums durch iteratives und inkrementelles Arbeiten sowie eine empirische Prozesssteuerung. Die Grundlagen dafür sind Transparenz über Beobachtungen und Erkenntnisse, die Reflektion über diese und das darauf basierende, kontinuierliche Anpassen des eigenen Handelns [26].

Diese Art zu arbeiten, erfordert kontinuierliche Beteiligung, eine offene Kommunikationskultur und Mut zur Veränderung. Gelingt diese vertrauensvolle Zusammenarbeit, können gesteigerte Qualität, motiviertere Mitarbeitende und eine schnellere und passgenauere Lösung für die Probleme der Nutzenden entstehen [27].

Neben den genannten Werten wie Offenheit und Mut tragen auch Prinzipien dazu bei, wie beispielsweise Entscheidungen zum spätesten sinnvollen Zeitpunkt zu treffen oder möglichst früh eine erste Version der Lösung zum Erproben durch die Nutzenden zu erstellen. Durch die regelmäßige Beurteilung des bereits Erreichten wird das Große und Ganze in kleinere Teile

zerlegt und man bleibt reaktionsfähig in Bezug auf sich ändernde Rahmenbedingungen oder Anforderungen [28].

Durch die Arbeit in kleinen und stabilen Teams wird ein gemeinsames Verständnis über die Ziele des Projekts und die Probleme der Nutzenden gefördert. Eine enge Zusammenarbeit aller Beteiligten – von der Entwicklung, über den Betrieb bis zur\* zum Nutzenden – ist essenziell für das Gelingen eines komplexen Projektes unter agilem Vorgehen [29].

Zu den bekannteren Vorgehensmodellen oder -frameworks gehören Scrum, Kanban und Extreme Programming (XP), wobei sich diese historisch teils gegenseitig beeinflusst haben und auch noch heute ergänzen können. Die Wahl eines Vorgehensmodells und insbesondere die Kombination ihrer Elemente sollte sehr bewusst durchgeführt werden.

Alle Modelle haben sich in der Praxis seit Jahrzehnten bewährt und ihr Bekanntheitsgrad erleichtert die Einarbeitung. Durch ihre Freiheitsgrade lassen sich alle Varianten an sich ändernde Anforderungen und Rahmenbedingungen anpassen.

Die markantesten Kennzeichen von Scrum sind der feste Iterationstakt (Sprint) und einige weitere Ereignisse (Review, Retrospektive) [30], die vorgegebenen Rollen und ihre Verantwortlichkeiten (Team, Product Owner, Scrum Master) sowie die Artefakte zur Steuerung der inhaltlichen Arbeit (Product Goal, Product Backlog, Sprint Backlog) und das nach jedem Sprint entstehende, fertige Inkrement. Scrum macht somit einige Vorgaben zur Sicherstellung eines empirischen Vorgehens und lässt doch gleichzeitig Spielräume für Product Management und weitere kontextabhängige Tätigkeiten [31].

Kanban setzt den stetigen Arbeitsfluss (Flow) und dessen Visualisierung (Kanban Board) in den Mittelpunkt und bietet Werkzeuge an, um diesen Flow zu optimieren (Work in Progress Limits). Ebenso wie in Scrum sind Feedbackschleifen für eine kontinuierliche Verbesserung zentral für einen empirischen Prozess. Im Gegensatz zu Scrum werden in Kanban auch die Rollen, Verantwortlichkeiten und bestehenden Prozesse zunächst als Startpunkt genutzt und mittels dieser Feedbackschleifen weiter entwickelt [32].

Extreme Programming (XP) beschreibt Prinzipien und Praktiken in den drei Bereichen Business bzw. Planung, Team sowie Technik. Dazu zählen beispielsweise die Kooperation zwischen Entwickelnden und Stakeholder bei der Planung, sowie der Gedanke, dass das gesamte Team für die erzeugten Artefakte Verantwortung trägt. Weiteres Kennzeichen sind bestimmte Methoden, wie die testgetriebene Entwicklung [33].

Für umwelt.info schlagen wir Scrum als grundlegendes Vorgehensmodell vor, welches durch Prinzipien und Elemente aus Kanban und XP angereichert wird. Zu den Scrum Ergänzungen zählen zum Beispiel das Kanban Board zur Visualisierung des Flows innerhalb eines Sprints oder die aus dem XP-Umfeld bekannten User Stories zur Formulierung von Anforderungen aus der Perspektive der Nutzenden [34].

Für diese Entscheidung wurden zwölf Anforderungen und Rahmenbedingungen geprüft und in Bezug zu den drei Vorgehensmodellen gesetzt. Dazu gehören zum Beispiel die Fragen nach der Anwendbarkeit auf Arbeitsschritte in den Ausbaustufen 1 – 4, wie erweiterbar und skalierbar ein Vorgehen ist oder wie hoch der Einarbeitungsaufwand einzuschätzen ist.

Im folgenden Abschnitt wird detaillierter beschrieben, wie die einzelnen Scrum-Elemente für die Entwicklung von umwelt.info angepasst bzw. durch welche Praktiken sie ergänzt werden. Diese Ergänzungen betreffen u.a. das Projektmanagement, das Zusammenspiel von verschiedenen Organisationseinheiten sowie konkrete Vorschläge für die Implementierung von Scrum (Rollen, Ereignisse, Artefakte) im Kontext dieses Projektes.

## 5.2 Anpassung an umwelt.info

Das Vorgehensmodell für die Entwicklung von umwelt.info beruht auf der Scrum-Methode, die jedoch für dieses Projekt angepasst wird. Eine detaillierte Beschreibung befindet sich im Anhang B.

Die Organisation besteht aus einem Team beim Nationalen Zentrum für Umwelt- und Naturschutzinformationen (umwelt.info), das mit inhaltlichen/redaktionellen Aufgaben sowie mit dem Betrieb des Systems betraut ist. Die inhaltliche Steuerung der Entwicklung liegt beim Chief Product Owner, der ebenfalls von umwelt.info gestellt wird (z. B. Leiter\*in des umwelt.info Teams). Somit ist sichergestellt, dass das Team von umwelt.info maximalen Einfluss auf den Ablauf der Entwicklung hat. Dies ist besonders wichtig, um unvollständig spezifizierte oder neu identifizierte Anforderungen möglichst effizient umzusetzen.

Der Chief Product Owner pflegt mit dem Product Backlog die zentrale Aufgabenliste, die in einer Abfolge von Sprints (sog. Gesamtteamsprints) abgearbeitet wird. Für einen oder mehrere dieser Sprints werden Verträge mit dem Auftragnehmer geschlossen, die als Meilensteine der Entwicklung angesehen werden können. Die eigentlichen Entwicklungsarbeiten werden von einem oder mehreren Teams geleistet. Die Anzahl der Teams bildet dabei einen Skalierungsfaktor, der sich mit der Zeit verändern kann. Neben dem Chief Product Owner gibt es eine\*n Projektleiter\*in, die für die organisatorische Abwicklung des Projekts zuständig ist und so die fachlich/inhaltlich geprägte Rolle des Chief Product Owners ergänzt.

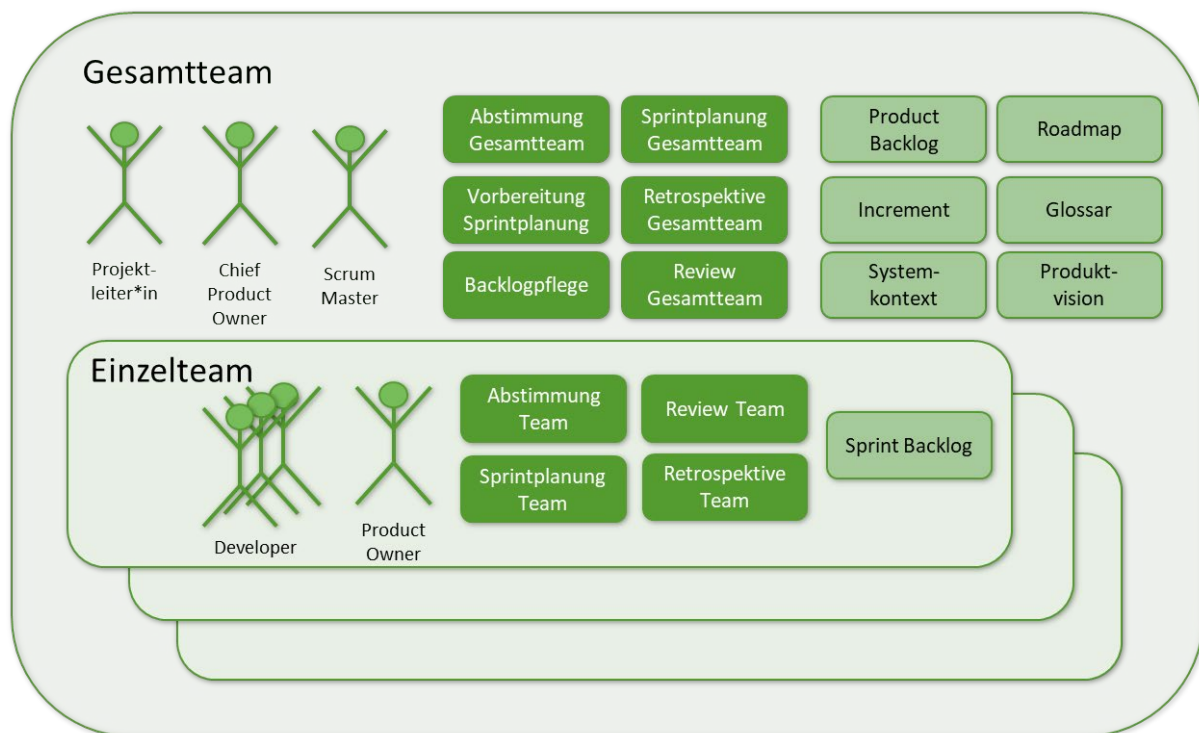
Die einzelnen Teams haben jeweils einen eigenen Product Owner, der die Kommunikation zum Chief Product Owner sicherstellt und diesen bei der Planung unterstützt. Auch in den einzelnen Entwicklungsteams finden die Arbeiten im Rahmen von Sprints statt. Diesen Sprints liegen Sprint-Backlogs zu Grunde, die aus dem zentralen Product Backlog abgeleitet werden. Bei der Zuteilung der Aufgaben an die einzelnen Teams wird der Chief Product Owner durch die Product Owner der einzelnen Entwicklungsteams unterstützt. Die Dauer der Einzelteamsprints und der Gesamtteamsprints können sich unterscheiden, wobei ein Gesamtteamsprint mehrere Einzelteamsprints umfassen kann. Ebenso kann sich ein Abruf über mehrere Gesamtteamsprints erstrecken. Aus dieser Trennung zwischen der Projektsteuerung durch den Chief Product Owner und der Projektdurchführung durch die einzelnen Teams ergibt sich eine Projektstruktur mit zwei Ebenen (vgl. Abbildung 8).

Die Projektsteuerung wird auf der Ebene des Gesamtteams durchgeführt und vor allem durch den Chief Product Owner und den Projektleitenden angetrieben. Die eigentliche Entwicklungsarbeit wird von den einzelnen Teams auf der Teamebene geleistet. Die Product Owner der einzelnen Teams und der Scrum Master verbinden die beiden Ebenen und unterstützen sowohl den Chief Product Owner als auch die Entwickelnden bei ihren Aufgaben.

Für beide Ebenen gibt es spezifische Ereignisse und mit den Backlogs auch spezifische Artefakte. Die meisten Artefakte jedoch sind in der Abbildung auf der Ebene des Gesamtteams angeordnet. Das drückt aus, dass diese Artefakte für alle Projektbeteiligten, unabhängig davon, welcher Ebene sie zugeordnet sind, direkte Relevanz besitzen.



**Abbildung 8: Verschiedene Ebenen in der Projektstruktur**



Quelle: Eigene Darstellung, con terra GmbH

### 5.3 Vertragsgestaltung

Ein agiles Vorgehen bei der Projektumsetzung erfordert die Schaffung geeigneter vertraglicher Grundlagen, die es dem UBA ermöglichen, Unterstützung für Beratungs- und Entwicklungsleistungen zu beschaffen. Daher wurde ein Vorschlag für die Vertragsgestaltung inklusive Preisgestaltung erarbeitet, der die oben beschriebene agile Vorgehensweise berücksichtigt.

Die agile Vorgehensweise ist insbesondere dadurch gekennzeichnet, dass noch keine vollumfängliche Beschreibung des Endergebnisses vorliegt, sondern diese erst im Verlaufe des Projekts konkretisiert wird. Ein detaillierterer Überblick der Problematik findet sich in [35]. Danach sind in der Vertragsgestaltung u. a. zu beachten, dass bereits zu Beginn des Projekts die Projektvision klar beschrieben wird, dass festgelegt ist, wer die Hauptverantwortung für den Entwicklungsprozess trägt und dass Dauer und Turnus der Sprints festgelegt sind.

Ebenso muss klar definiert sein, welche Aufgaben und Rollen durch den Auftraggeber wahrgenommen werden und wo die Verantwortung für welche Teile der Dokumentation liegt. Darüber hinaus sollten im Vertrag Details zu Kündigungsrechten, zu Modalitäten bezüglich Abnahme und Gewährleistung geregelt sein. Schließlich ist die Festlegung von Nutzungsrechten am gesamten System oder an Teilen davon ein wichtiger Bestandteil der vertraglichen Grundlagen.

In Abstimmung mit dem UBA wurde ein Vertragsmuster für eine Rahmenvereinbarung entworfen. Darin werden alle Regelungen getroffen, die über das gesamte Projekt gelten. Über die Rahmenvereinbarung werden Einzelabrufe vereinbart, für die jeweils ein EVB-IT-Vertrag abgeschlossen wird. Bei den EVB-IT-Verträgen sind unterschiedliche Typen möglich (z. B. EVB-IT-Erstellung, EVB-IT-Dienstleistung, EVB-IT-Überlassung). Einzelabrufe werden entweder zum Festpreis oder nach Aufwand (mit Obergrenze) abgeschlossen. Die Einzelabrufe sind eigen-

ständige Verträge und, im Falle von Werkverträgen, kann es Abnahmen geben. Ein Einzelabruf kann einen oder mehrere Sprints umfassen.

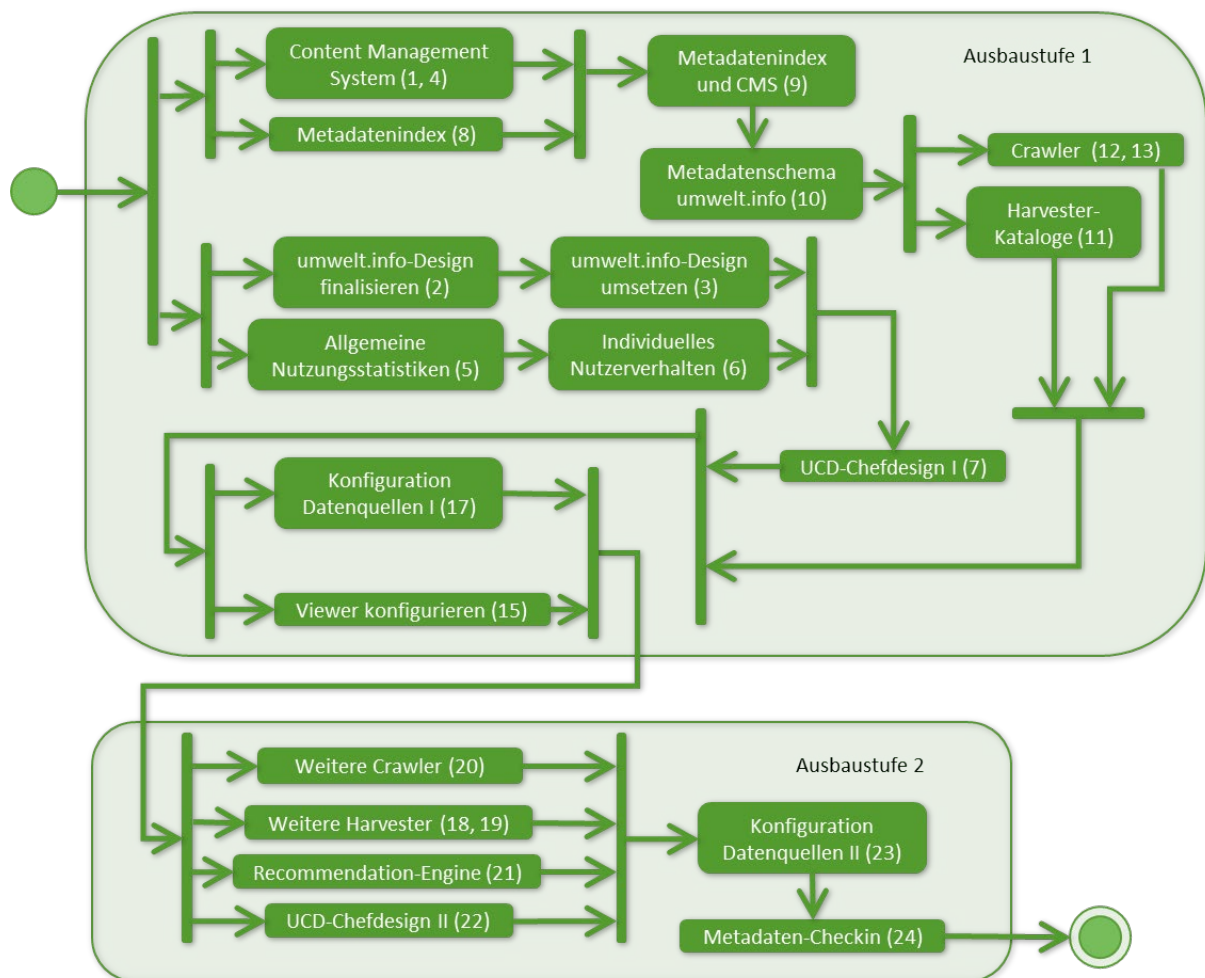
Die Abfolge der Einzelabrufe kann auch auf der Roadmap abgebildet und bereits zu Beginn des Vorhabens festgelegt werden. Dies kann zunächst auch geschehen, ohne für alle Meilensteine feste Termine vorzusehen.

## 5.4 Planung der Stufen 1 und 2 von umwelt.info

Für die Planung der ersten Ausbaustufen von umwelt.info wurden erste Entwürfe der Roadmap und des Backlogs erstellt. Ein Entwurf der Produktvision ist bereits in Kapitel 2.1 formuliert. Damit kann die Realisierung nach dem beschriebenen agilen Vorgehen begonnen werden.

Auf Basis der Aktivitäten, die im Rahmen der Machbarkeitsstudie [1] für die dort formulierten Ausbaustufen 1 und 2 identifiziert wurden sowie der erstellten Systemarchitektur, wurde der Entwurf einer Roadmap erstellt. Diese Roadmap zeigt eine mögliche Reihenfolge von Arbeitspaketen auf, auf deren Grundlage die Definition der Sprints erfolgen kann (vgl. Abbildung 9). Zu Beginn des Entwicklungsprojekts muss diese Roadmap nochmals verifiziert und ggf. an geänderte Rahmenbedingungen oder Anforderungen angepasst werden.

**Abbildung 9: Entwurf einer Roadmap**



Quelle: Eigene Darstellung, con terra GmbH



Die Aktivitäten in der Roadmap wurden im Vergleich zur Machbarkeitsstudie angepasst und weichen in ihrer Benennung teilweise ab. Der Verweis auf die jeweiligen Aktivitäten dort erfolgt mittels der Nummer, die zur Aktivität eingetragen ist. Die Roadmap zeigt keine lineare Abfolge von zu entwickelnden Modulen. Vielmehr ergeben sich zahlreiche Möglichkeiten der parallelen Bearbeitung durch verschiedene Teams. Gelingt eine solche Parallelisierung der Entwicklung, lässt sich die Zeitdauer der Realisierung deutlich verringern.

Das Product-Backlog kann aus den Aktivitäten der Machbarkeitsstudie [1], dem Designkonzept sowie der Systemarchitektur abgeleitet werden. Nach dem Formulieren von User Stories und dem Bilden von Clustern aus diesen User Stories wurden inhaltliche Themen (sog. Epics) identifiziert und den geclusterten User Stories in Form einer Kreuztabelle in einer n:m-Beziehung zugeordnet. Die einzelnen Teilaufgaben, die für die Umsetzung der Epics durchgeführt werden, bilden dann die verschiedenen Einträge im Product-Backlog. Der Backlogentwurf liegt dem UBA vor und kann dort auf Anfrage eingesehen werden. Hier wird nur die Struktur der einzelnen Einträge beschrieben.

Das Product Backlog ist aus Items aufgebaut. Jedes Item beschreibt dabei eine Aufgabe, die im Rahmen des Projekts durchgeführt werden muss. Jede Aufgabe verfügt über bestimmte Informationen. Diese liegen bei der Erstellung des Items teilweise noch gar nicht oder nur unvollständig vor und werden erst im Verlauf der Backlogpflege ergänzt oder verfeinert. Dies ist die Aufgabe des Chief Product Owners, bei der er durch die Product Owner der einzelnen Teams unterstützt wird. Jedes Item verfügt über folgende Informationen:

- ▶ **Identifikationsnummer**  
Eine eindeutige numerische ID, mit der die Aufgabe referenziert werden kann.
- ▶ **Name**  
Ein möglichst aussagekräftiger Name für die Aufgabe.
- ▶ **Referenz auf Machbarkeitsstudie (optional)**  
Nummer der Aktivität aus der Machbarkeitsstudio, sofern sich das Item einer solchen Aktivität zuordnen lässt.
- ▶ **Ausbaustufe**  
Ausbaustufe aus der Machbarkeitsstudie. Ein Item kann in mehreren Ausbaustufen oder auch in allen relevant sein. In diesem Fall leiten sich aus einem Item wiederholt Aufgaben für Sprints ab.
- ▶ **Beschreibung**  
Inhaltliche Beschreibung der Aufgabe, auch mit technischen Anmerkungen. Diese Beschreibung wird im Laufe des Projekts und mit zunehmender Reife des Eintrags mit weiteren Details, Entscheidungen, Fragen und Anmerkungen versehen.
- ▶ **Priorität (optional)**  
Eigentlich entspricht die Reihenfolge der Backlogseinträge auch der Priorität der Items. Insbesondere bei vielen Einträgen, die nicht kurzfristig realisiert werden sollen, oder bei Einträgen, die noch den Charakter von bloßen Ideen haben, ist es sinnvoll, mit Hilfe dieses Attributs explizit die Bedeutung für das Projekt zu kennzeichnen.
- ▶ **Verbundene Items (optional)**  
Hierbei handelt es sich um eine Liste von anderen Items, die zum aktuellen in einer Beziehung stehen. Das können Aufgaben sein, die zwingend erledigt sein müssen, bevor mit der aktuellen begonnen werden kann, Aufgaben, die während der Arbeit an der aktuellen

besonders berücksichtigt werden sollen oder auch doppelte Einträge. Wenn eine komplexe Aufgabe in mehrere einzelnen Aufgaben aufgeteilt wird, dann ist die Verbindung der neuen Einträge zum Ursprungsitem auch in diesem Attribut zu speichern.

- ▶ **Vorbedingungen (optional)**  
Hier können Rahmenbedingungen formuliert werden, die für die Realisierung der aktuellen Aufgabe erfüllt sein müssen.
- ▶ **Akzeptanzkriterien**  
Hier werden die Kriterien festgelegt, die für die Entscheidung über die Erfüllung der Aufgabe herangezogen werden. Die Akzeptanzkriterien müssen formuliert sein, wenn ein Item zur Realisierung in einen Sprint aufgenommen wird. Dies ist die Zuständigkeit des Chief Product Owners.
- ▶ **Aufwand**  
Dieses Attribut enthält den geschätzten Aufwand. Es muss gefüllt sein, bevor die Entscheidung über die Realisierung des Eintrags gefällt wird.
- ▶ **Komponenten**  
Dies sind die betroffenen Teile der Systemarchitektur. Sind die Entwicklungsteams analog zur Architektur strukturiert, so beeinflusst diese Information die Zuweisung einer Aufgabe zu einem bestimmten Team.
- ▶ **Verbundene Dokumente**  
Ein Backlogeintrag kann zur weiteren Information mit verschiedenen Dokumenten versehen werden. Dies können Entwürfe der Benutzeroberfläche aus der Machbarkeitsstudie sein, Schnittstellendefinitionen, Wertelisten u.v.a.

Beim Erstellen eines neuen Eintrags ist die Aufgabe in der Regel noch sehr grob definiert und nur überblickshaft beschrieben. Wichtige Informationen wie z. B. die Aufwandsschätzung liegen noch nicht vor. Im Laufe der Zeit wird das Item vervollständigt. Dabei werden neue Items abgeleitet, sodass kleinere Aufgaben entstehen. Neben diesem Mittel zur Teilung komplexer Aufgaben sieht der Entwurf des Product-Backlogs noch vor, dass eine Aufgabe innerhalb eines Items in Teilaufgaben (Subitems) untergliedert werden kann.

## 6 Organisationskonzept

Das umwelt.info-Portal als zentrale Stelle für Verwaltung und Verfügbarmachung von beschreibenden Metadaten zu Umweltdaten und Umweltinformationen unterschiedlicher Herkunft ist in eine umfangreiche und heterogene Infrastruktur von Anbietern und Nutzenden eingebunden. Der Organisationseinheit umwelt.info als zuständiger Stelle für Realisierung und Betrieb dieses Portals sind dementsprechend vielfältige Aufgaben zugewiesen.

In diesem Kapitel wird die Organisation von umwelt.info beschrieben. Dabei werden die Aufgabenbereiche benannt und Organisationsuntereinheiten zugewiesen.

Bei dieser Beschreibung wird berücksichtigt, dass zwischen der Entwicklungs- und der Betriebsphase des umwelt.info-Systems unterschieden werden muss, auch wenn diese Phasen nicht vollständig voneinander getrennt werden können. Bei fortschreitendem Übergang von der Entwicklungs- zur Betriebsphase verschieben sich so die Kernaufgaben der Organisationseinheit und damit auch die Gewichte und Personalbedarfe der einzelnen Untereinheiten.

In diesem Kapitel wird zunächst die Einbettung der Organisationseinheit umwelt.info in die umgebende Struktur aus Anbietern und Nutzenden beschrieben. Dabei werden Anforderungen aufgeführt, die sich während des Betriebs aus dem Verhältnis der einzelnen Strukturkomponenten zu umwelt.info ergeben (vgl. Kap. 6.1). Im Kapitel 6.2, das die Organisationseinheit selbst beschreibt, werden die verschiedenen Aufgaben benannt, die sich aus der Entwicklung nach dem Vorgehensmodell (vgl. Anhang B) und dem Betrieb in der zuvor beschriebenen Anbieter- und Nutzungsstruktur ergeben.

### 6.1 Beziehungen zu Anbietenden und Nutzenden

Das System umwelt.info als Metainformationssystem und Metadatenportal fungiert als Mittler zwischen Informationsanbietern und Informationsnutzern. Bestimmte Organisationen können dabei auch in einer Doppelrolle sowohl als Anbieter als auch als Nutzende auftreten. Grundlegende Aufgabe des Systems ist es also, einerseits Informationsnutzer sowohl bei der Suche nach Informationen zu unterstützen als auch beim Zugriff auf gefundene Datenbestände. Andererseits soll umwelt.info den Anbietern von Umweltinformationen eine Plattform bieten, auf der sie ihre Inhalte an potenzielle Nutzende vermitteln können.

umwelt.info besitzt einen Datenbestand, in dem im Wesentlichen Metadaten gespeichert sind. Das sind Daten, die andere Datenbestände, sog. Primärdaten, beschreiben. Diese Primärdaten werden von Informationsanbietern beschafft, verwaltet und gepflegt. Die Informationen über diese Daten werden vom jeweiligen Informationsanbieter bereitgestellt, damit sie in den Metadatenbestand von umwelt.info integriert werden können. Dabei ist es möglich, dass die Bereitstellung der Metadaten innerhalb von umwelt.info aktiv auf Wunsch des Datenanbieters erfolgt oder durch Akquise von Seiten umwelt.info.

Neben den Metadaten umfasst umwelt.info auch redaktionelle Beiträge. Diese beleuchten bestimmte umweltrelevante Themen und bieten Informationen und Hilfestellungen zur Anwendung des Portals an. Auch die angestrebte Förderung der Kommunikation und des fachlichen Austauschs zu Umweltinformationen kann durch Formate wie Veranstaltungskalender, Newsletter oder ein Nutzerforum erreicht werden.

Informationsnutzende durchsuchen die Inhalte von umwelt.info nach bestimmten Themen und finden dabei Informationen über die Datenbestände der Anbieter, sowie über mögliche Zugriffswege auf die Primärdaten selbst. Weitere mögliche Ergebnisse dieser Suche sind die redaktionell aufbereiteten Angebote von umwelt.info.

Für die Kommunikation zwischen umwelt.info, den datenhaltenden Stellen sowie bestimmten Kernnutzern wird eine Gremienstruktur ähnlich dem bisherigen Projektbeirat fortgeführt. Insbesondere bei Kontakt zu den Ministerien der Länder und strategischen Entscheidungen wird das BMUV eng mit einbezogen.

Aus den Beziehungen zwischen den verschiedenen Komponenten der Gesamtstruktur leiten sich unterschiedliche Anforderungen an die Organisationseinheit umwelt.info ab, die im Folgenden aufgeführt werden.

### 6.1.1 Beteiligung von Informationsanbietern

Es gibt Informationsanbieter unterschiedlichen Typs. Diese Typen ergeben sich zum einen aus den verschiedenen Arten von Informationen, die bereitgestellt werden: Strukturierte Daten in Datenbanken oder definierten Dateiformaten, Dienste (z. B. Geodatendienste), Dokumente (Text, Tabellen, Abbildungen, ...), Webseiten, diverse Dateien und Dateisammlungen.

Außerdem gibt es Unterschiede bei der Art der Bereitstellung. Häufig stellen Datenanbieter bereits Metadaten zu ihren Primärdaten oder Diensten zur Verfügung. Daneben gibt es jedoch auch Organisationen, die nur Primärdaten verwalten, ohne dass beschreibende Metadaten existieren. Schließlich gibt es auch Informationsanbieter, die ausschließlich Metadaten zur Verfügung stellen. Gerade unter den ersten beiden Gruppen gibt es Anbieter, die auch Beratung und Hilfe bei der Bereitstellung von Metadaten und ggf. auch Primärdaten wünschen.

Die Qualität in Hinblick auf Vollständigkeit, Genauigkeit, Auflösung und Aktualität ist bei verschiedenen Informationsanbietern unterschiedlich. Zuverlässige Informationen zur Metadatenqualität sind nicht immer verfügbar. Dies hat Einfluss auf die Nutzbarkeit der Daten und muss daher aus den Metadaten ablesbar sein. Es ist also notwendig, dass umwelt.info die Qualität der Metadaten beurteilt.

Ziel ist es, zunächst alle Angebote in umwelt.info verfügbar zu machen, die bereits im Internet verfügbar und automatisiert über Harvester und Crawler verarbeitet werden können. Bei allen weiteren Daten und Informationen sollte zunächst eine gezielte Beratung zu den technischen Anforderungen erfolgen. In Zusammenarbeit mit den datenhaltenden Stellen wird umwelt.info die Metadatenintegration kuratieren und die Metadatenqualität ggf. optimieren. Eventuell könnte ein Label „umwelt.info Unterstützer“ helfen, um datenhaltende Stellen zu motivieren ihre Daten in bester Qualität zu referenzieren und verfügbar zu machen.

Aus den unterschiedlichen Typen von Informationsanbietern leiten sich folgende Anforderungen an die Organisationseinheit umwelt.info ab:

- ▶ Bereitstellen von Schnittstellen, mit denen vorhandene Metadaten aktiver Informationsanbieter in umwelt.info integriert werden können.
- ▶ Bereitstellen von Erfassungsfunktionalität, die es Informationsanbietern ermöglicht, Metadaten zu ihren Primärdaten innerhalb von umwelt.info zu erstellen.
- ▶ Automatisiertes Importieren von Metadaten von Anbietern, die diese nicht selbst zur Verfügung stellen.
- ▶ Automatisiertes Erstellen von Metadaten zu vorhandenen Primärdaten von Informationsanbietern, die ihre Metadaten nicht selbst in umwelt.info einstellen.
- ▶ Beratung von potenziellen Informationsanbietern in Bezug auf die Erstellung und das Publizieren von Metadaten.

- ▶ Information und Beratung von Informationsanbietern bezüglich aktueller technischer Standards und Spezifikationen.
- ▶ Initiale und laufende Qualitätssicherung und Aktualisierung von Metadaten innerhalb von umwelt.info.
- ▶ Organisation des Austauschs zu bestehenden und in Entwicklung befindlichen Angeboten mit grundsätzlichem Mehrwert für Umwelt und Natur in den relevanten Bundes- und Landesbehörden, Kommunen sowie in der Wirtschaft und Wissenschaft.
- ▶ Durchführen von Info-Veranstaltungen. Mit Hilfe klassischer Informationsveranstaltungen und Konferenzen können ebenso wie über Webmeetings die Anforderungen, die Informationsanbieter erfüllen sollen, vermittelt werden. Weitere Beteiligungsformate stellen geschlossene Arbeitskreise der Anwendenden oder Arbeitsgruppen mit definierten Themen und festgelegtem Teilnehmerkreis sowie Hackathons oder Datathons dar.
- ▶ Durchführen von Online-Konsultationen, um Meinungen aus dem Kreis der informationshaltenden Stellen einzuholen.
- ▶ Einrichten eines User-Forums als Kommunikationsportal zum Austausch innerhalb des Kreises der Datenhaltenden Stellen und Nutzenden von umwelt.info über Projekte und Lösungen.
- ▶ Erstellen einer Dokumentation mit genauen Handlungsanweisungen. Diese hilft den informationshaltenden Stellen bei der Einbindung ihrer Angebote in umwelt.info. Online-Tutorials können eine solche Dokumentation ergänzen.
- ▶ Beratung und Schulung: umwelt.info sollte gezielte Beratungspakete und Schulungsangebote konzipieren und durchführen.
- ▶ Einrichten einer festen Kontaktstelle für Informationsanbieter, die schnell und individuell Hilfestellung bei Fragen zu Themen der Daten- und Metadatenbereitstellung bieten kann.
- ▶ Präsenz auf Veranstaltungen anderer Träger wie Messen, Tagungen etc.

### **6.1.2 Beteiligung von Informationskonsumenten**

Auch die Profile der Nutzenden von umwelt.info sind sehr unterschiedlich. Zum einen gibt es sehr intensiv Nutzende mit hohen Ansprüchen an die Aussagekraft der verwalteten Metadaten und komplexen Anforderungen an Such- und Präsentationsfunktionen. Zum anderen wird es auch gelegentlich Nutzende geben, für die der einfache, intuitive Zugang zu Metadaten und weiterführend auch zu den beschriebenen Primärdaten wichtig ist.

Neben Metadaten zur Beschreibung von Primärdatenbeständen verschiedener Informationsanbieter werden in umwelt.info auch redaktionell aufbereitete Inhalte zu ausgewählten Themen den Anwendenden bereitgestellt.

Insgesamt leiten sich aus den Beziehungen von umwelt.info zu den Informationskonsumenten folgende Anforderungen ab:

- ▶ Bereitstellen einer Benutzeroberfläche, mit deren Hilfe es möglich ist, leicht im Metadatenbestand zu navigieren,
- ▶ Bereitstellen mächtiger Suchfunktionen, um bestimmte Themen schnell zu finden,

- ▶ Bereitstellen einer Hilfefunktion für unerfahrene Benutzer und von beispielhaften Nutzungen z. B. durch Data Stories o.ä.,
- ▶ Publizieren von Informationen zu bestimmten Themen,
- ▶ Bekanntmachen des Angebots von umwelt.info für Nutzende unterschiedlichen Typs,
- ▶ Bearbeitung und Abwägung von Anforderungen an Bund, Länder, Kommunen, Wirtschaft und Wissenschaft als informationshaltende Stellen,
- ▶ Erstellen eines Beteiligungsportals, das Nutzenden die Möglichkeit bietet, direkt Vorschläge oder Anmerkungen zu umwelt.info zu äußern und das zudem Kommunikation und Austausch in der Nutzungsgemeinschaft fördert. Dazu gehört auch die Moderation dieses Portals, also die schnelle und angemessene Reaktion auf Beiträge (Beantworten von Fragen, Richtigstellen von Falschinformationen u. a.).
- ▶ Einbinden von Informationskonsumenten mit Hilfe offener Beteiligungsformate wie Hackathons, Anwenderkreisen oder öffentlichen Tagungen.
- ▶ Analysieren der Nutzungsintensität durch regelmäßige Evaluation von Nutzungsstatistiken und Ableiten von Handlungsempfehlungen aus den Ergebnissen.
- ▶ Präsenz auf öffentlichen Veranstaltungen anderer Organisationen wie Messen, Tagungen etc.
- ▶ Bereitstellen von Schnittstellen mit denen in umwelt.info indizierte Metadaten abgefragt werden können.

## **6.2 Organisationseinheit umwelt.info**

Als Betreiber des umwelt.info-Portals ist das neu eingerichtete Nationale Zentrum für Umwelt- und Naturschutzinformationen mit Sitz in Merseburg vorgesehen [4]. An dieser Stelle werden Aufgaben und Struktur im Hinblick auf Realisierung und Betrieb des Portals beschrieben. Diese Aufgaben leiten sich aus den Anforderungen des entworfenen Vorgehensmodells für die Entwicklung (vgl. Anhang B) und aus den Beziehungen zu Informationslieferanten, Nutzenden und BMUV als Dienstaufsicht während des Betriebs ab (vgl. Kap. 6.1). Die identifizierten Aufgaben werden in die Aufgabenpakete der Entwicklung, des Betriebs sowie übergreifende Aufgaben untergliedert.

Entwicklung und Betrieb von umwelt.info lassen sich nicht vollständig trennen. Zwar fallen zunächst fast ausschließlich Aufgaben aus dem Bereich der Entwicklung an. Diese werden in Kap. 6.2.1 dargestellt. Durch den fließenden Übergang hin zur Betriebsphase wird der Anteil der Aufgaben für den operationellen Betrieb aber immer größer, wobei auf Grund neuer Anforderungen an umwelt.info die Entwicklungsaufgaben nie ganz entfallen werden. Aufgaben für die Betriebsphase werden in Kap. 6.2.2 beschrieben.

Begleitet werden beide Phasen von übergreifenden Aufgaben, die die Verwaltung der Organisationseinheit umwelt.info und deren allgemeine Präsentation nach außen betreffen. Diese Aufgaben werden in Kap. 6.2.3 behandelt.

### **6.2.1 Entwicklung des Portals umwelt.info**

Nach dem für die Entwicklung von umwelt.info entworfenen Vorgehensmodell erfolgt die Realisierung von umwelt.info in enger Kooperation zwischen der Organisationseinheit



umwelt.info und einem oder mehreren Auftragnehmern. umwelt.info obliegen dabei im Wesentlichen folgende Aufgaben:

- ▶ E1 - Steuerung der Entwicklung: Unter die Steuerung der Entwicklung fallen die Koordination der verschiedenen Entwicklungsteams und die Koordinierung der externen Beratenden und Entwickelnden. Ebenso umfasst die Aufgabe die Planung und Durchführung funktionaler und nicht-funktionaler Tests.
- ▶ E2 - Spezifikation von Anforderungen: Im Rahmen der Spezifikation von Anforderungen müssen Anforderungen durch den Kontakt mit potenziellen Nutzenden ermittelt werden. Des Weiteren müssen sie im Lauf der Entwicklung schrittweise konkretisiert und priorisiert werden. Aktuelle Standards und Spezifikationen, neue Verfahren zur Daten- und Informationsrecherche und -aufbereitung oder Entwicklungen im Bereich KI und Linked Data sollten erkannt und auf ihre Eignung und Validität für umwelt.info hin geprüft werden, um ggf. in neue Anforderungen einzufließen.
- ▶ E3 - Kommunikation und Interaktion mit datenhaltenden Stellen: Im Vorgehenskonzept für die Entwicklung von umwelt.info wurde bereits verdeutlicht, dass der Betrieb des umwelt.info-Systems nicht komplett von der Entwicklung getrennt ist, sondern dass ein fließender Übergang zwischen diesen beiden Phasen existiert (vgl. Anhang B). Darüber hinaus sind Ideen, Anforderungen und Kenntnisse der verschiedenen datenhaltenden Stellen bereits bei der Entwicklung von großem Nutzen. Insbesondere Informationen über vorhandene Daten und Metadaten sind eine wichtige Information und sollten bereits in der Entwicklungsphase berücksichtigt werden. Aus diesen Gründen sollte die Einbindung verschiedener Datenanbieter möglichst früh als Aufgabe angesehen und bei der Kapazitätsplanung berücksichtigt werden. Beispiele für Ergebnisse der Kommunikation mit datenhaltenden Stellen sind Projekte, um (Meta-)Daten fachlich oder technisch zu harmonisieren oder die koordinierte Neuentwicklung von Datenbanken, oder Informationssammlungen durch datenhaltende Stellen. Diese Aufgaben bleiben während des Betriebs vollumfänglich erhalten.
- ▶ E4 - Entwicklung und Konfiguration einzelner Komponenten: Das Team von umwelt.info kann auch einzelne Komponenten entwickeln und/oder konfigurieren. Hierfür wird ein entsprechendes IT-technisches Wissen benötigt.

### 6.2.2 Betrieb des Portals umwelt.info

Beim Betrieb des Portals umwelt.info gibt es technische, inhaltliche und administrative Aspekte zu berücksichtigen. Die technischen Aspekte betreffen dabei den eigentlichen Betrieb der Software.

- ▶ B1 - Technischer Betrieb: Der technische Betrieb umfasst die Bereitstellung, Wartung und Pflege der Komponenten von umwelt.info sowie den technischen Support für diese Komponenten. Darüber hinaus ist auch die Kontrolle des Betriebs, also die Analyse der Nutzungsintensität und die Erstellung von Nutzungsstatistiken etc. Teil des technischen Betriebs. Außerdem gehören zu diesem Aufgabenbereich die Koordinierung und die Kontrolle von Dienstleistern, sofern der Betrieb durch externe Partner durchgeführt wird. Die inhaltlichen Betriebsaspekte sind durch den Aufbau und die Pflege des umwelt.info-Metadatenbestandes geprägt. Außerdem umfasst dieser Bereich Aufgaben im Kontakt mit (potenziellen) Informationsanbietern und Nutzenden sowie der Öffentlichkeit.

- ▶ B2 - Erstellen von Metadatensätzen: Durch Import vorhandener Metadaten oder durch automatisierte Erstellung werden beschreibende Metadaten erstellt. Im Einzelfall kann diese auch durch manuelle Eingabe erfolgen. Zur Erstellung von Metadatensätzen gehört ggf. auch die Ergänzung von Metadaten durch bestimmte zusätzliche Informationen wie Vertrauenswürdigkeit, Aktualität, Relevanz, Rechtssicherheit etc.).
- ▶ B3 - Qualitätssicherung der Metadaten: Diese Aufgabe umfasst die Prüfung der Qualität von eingegebenen, generierten oder importierten Metadaten in Hinsicht auf ihre Korrektheit und Aktualität. Außerdem gehört dazu die Suche nach doppelten Einträgen und fehlerhaften Verweisen auf Primärdaten.
- ▶ B4 - Qualitätssicherung der Funktionalität: Neben der Qualität der Metadaten in umwelt.info muss auch die Funktionalität geprüft werden. Diese Prüfung hat dabei nicht den technischen Fokus der Prüfung der Lauffähigkeit. Vielmehr werden inhaltliche Aspekte geprüft wie z. B. die Qualität, Vollständigkeit und Aussagekraft von Suchergebnissen, Schlagwortzuordnungen etc.
- ▶ B5 - Erstellen von Fachinhalten: Auf der Basis von Fachinformationen, die in umwelt.info bekannt und beschrieben sind, sollen zu ausgewählten Themen, die z. B. besondere aktuelle Dringlichkeit besitzen, fachliche redaktionelle Beiträge und Formate erstellt werden. Dazu werden Informationen aus verschiedenen Quellen anschaulich und auf vielfältige, zielgruppenorientierte Weise aufbereitet. Solche Beiträge können dann ihrerseits auch durch neue Metadaten beschrieben werden, die in umwelt.info eingepflegt werden und somit durch die Suche im Metadatenbestand auffindbar sind. Dazu gehören Attribute wie Verfasser, Zeitpunkt der Erstellung, Schlagworte etc.  
Das Konzipieren und Erstellen dieser Beiträge und Formate und der beschreibenden Metadaten ist Inhalt dieser Aufgabe.
- ▶ B6 - Beratung von Informationsanbietern: Zu dieser Aufgabe gehören beispielsweise das Aufzeigen von Wegen für die Erstellung von Metadaten und für deren Bereitstellung durch umwelt.info, die Funktion als direkte Kommunikationsstelle zwischen Informationsanbietern und umwelt.info, die Organisation eines Arbeitskreises für Informationsanbieter, die Organisation von Arbeitsgruppen zu technischen Themen (z. B. Metadatenformate, Schnittstellen) und die Durchführung von Fortbildungsveranstaltungen zu unterschiedlichen Themen wie Technologie, Standards, Qualität usw.
- ▶ B7 - Gremien- und Öffentlichkeitsarbeit: In diesen Aufgabenbereich fallen z. B. das Akquirieren neuer Informationsanbieter und Konsumenten, die Mitwirkung in Gremien wie dem Open Government Partnership, dem Steuerungskreis Digitalisierung der Bund-Länder-Arbeitsgruppe Umwelt und Digitalisierung oder der Geodateninfrastruktur Deutschland, die Durchführung von Hackathons und die Einbeziehung von Nutzenden durch die Organisation von Tagungen oder Arbeitskreisen von Nutzenden. Auch die Pflege eines Userforums und der Aufbau einer Gemeinschaft der Nutzenden gehören in diesen Aufgabenbereich (vgl. auch Kap. 6.1.2).
- ▶ B8 - Nutzerverwaltung: diese Verwaltungsaufgabe umfasst die Bearbeitung von Registrierungsanträgen, die Verwaltung des Verzeichnisses der registrierten Nutzenden und die Vergabe von Zugriffsrechten.



### 6.2.3 Übergreifende Aufgaben

Die übergreifenden Aufgaben bestehen aus der Leitung und Administration der Organisationseinheit und aus abteilungsübergreifenden organisatorischen Aufgaben. Zu diesen gehört auch die technische Durchführung von Veranstaltungen.

- ▶ Ü1 - Leitung der Organisationseinheit
- ▶ Ü2 - Organisation und Planung: Das umfasst auch die organisatorische Zusammenarbeit und Kommunikation mit Ländern und Kommunen sofern es sich dabei nicht um konkret fachliche Arbeiten handelt. Technische und fachlich/inhaltliche Aspekte sowie die konkrete Zusammenarbeit mit Datenanbietern fallen in den Aufgabenbereich B6.
- ▶ Ü3 - Veranstaltungsmanagement
- ▶ Ü4 - Juristische Unterstützung: In Hinblick auf die juristische Unterstützung ist derzeit nicht klar, ob es tatsächlich notwendig ist, entsprechende Kapazitäten innerhalb der Organisationsstruktur umwelt.info vorzuhalten oder ob diese Aufgaben besser durch Juristen der tragenden Institutionen oder durch externe Beratende abgedeckt werden können. Dennoch ist Bedarf an juristischer Unterstützung z. B. bei der Klärung für lizenzrechtlicher Fragen oder in Bezug auf Fragen des Datenschutzes nötig.
- ▶ Ü5 - Controlling und Vertragsmanagement: Diese Aufgabe umfasst die fortlaufende Kontrolle des Budgets und der Vertragserfüllung bzw. die Steuerung notwendiger Anpassungen.
- ▶ Ü6 - Projektdokumentation und Abstimmung mit politischen Zielen des BMUV.

## 7 Aktualisierter Rechtsrahmen zur Umsetzung

Die rechtlichen Rahmenbedingungen für den Aufbau eines Umwelt- und Naturschutzinformationssystems wurde in der UNIS-D Machbarkeitsstudie [1] bereits ausführlich dargestellt. Im Folgenden werden Änderungen im Rechtsrahmen nach Abschluss der Machbarkeitsstudie kurz dargestellt und eingeordnet. Im Anschluss werden die rechtlichen Anforderungen, die für das inzwischen weiter konkretisierte Umsetzungskonzept des Portals relevant sind, kurz überblicksartig dargestellt und daraus Handlungsempfehlungen abgeleitet.

### 7.1 Neuregelungen seit Erstellung der Machbarkeitsstudie

#### 7.1.1 Umweltinformationsgesetz (UIG)

Mit der UIG-Novelle vom 3. März 2021 wurde die Zuständigkeit des Beauftragten für Informationsfreiheit auf Umweltinformationen und Anträge nach UIG erweitert (§ 7a UIG). Weiterhin wurde ein neuer § 10 Abs. 3 Satz 3 UIG eingefügt, der festlegt, dass bestimmte Umweltinformationen, die aktiv verbreitet werden müssen, über das UVP-Portal des Bundes bereitgestellt werden können. Konkret betrifft dies Zulassungsentscheidungen sowie die zusammenfassende Darstellung und Bewertung von Umweltauswirkungen nach §§ 24, 25 UVPG. In § 10 Abs. 8 UIG wurden zudem neue Verordnungsermächtigungen zur UVP-Portalverordnung (UVP-PortV) sowie zur Aktualisierung von veröffentlichungspflichtigen Umweltinformationen eingefügt.

Die Neuregelungen sind für die Umsetzung von umwelt.info insgesamt von untergeordneter Bedeutung. Die neue Rolle des Informationsbeauftragten ist für den Portalbetrieb wenig relevant, und ggf. veröffentlichungspflichtige Umweltinformationen werden von umwelt.info größtenteils nicht selbst gehalten, sondern lediglich verlinkt bzw. auffindbar gemacht (siehe Kap. 7.2.1).

#### 7.1.2 Urhebergesetz (UrhG)

Die Novelle des UrhG vom 31. Mai 2021 [34] zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarkts ist für die Umsetzung von umwelt.info nicht relevant. Insbesondere bleibt es dabei, dass keine Haftung für bloß verlinkte Inhalte besteht (siehe auch Kap. 7.2.1).

#### 7.1.3 Telekommunikations-Telemedien-Datenschutz-Gesetz (TTDSG)

Das TTDSG wurde im Sommer 2021 erlassen, um die datenschutzrechtlichen Vorgaben für Telekommunikations- und Telemediendienste an die Bestimmungen der Datenschutz-Grundverordnung (DSGVO) anzupassen. Für umwelt.info (einen Telemediendienst) ist hier vor allem die nun explizit geregelte Verpflichtung relevant, eine Einwilligung für das Setzen nicht unbedingt erforderlicher Cookies einzuholen (§ 24 TTDSG).

#### 7.1.4 Geologiedatengesetz (GeolDG)

Das GeolDG, das bei Fertigstellung der Machbarkeitsstudie erst im Entwurf vorlag, ist zwischenzeitlich erlassen. Es regelt die Übermittlung, die dauerhafte Sicherung und öffentliche Bereitstellung dieser Daten sowie ihre Zurverfügungstellung zur Erfüllung öffentlicher Aufgaben (§ 1 GeolDG).

Wie im Rahmen der Machbarkeitsstudie erwähnt, regelt das Gesetz den Umgang mit Geologiedaten bundesweit einheitlich und macht hierbei umfassende Vorgaben für die

Veröffentlichung dieser Daten aber auch für ihre (vorgelagerte) Übermittlung durch Dritte, einschließlich Behörden (§§ 8 ff. GeolDG). Die zuständigen (Landes-)Behörden sind dann für die öffentliche Bereitstellung der Daten verantwortlich. Für die Modalitäten der Veröffentlichung verweist § 5 Abs. 3 GeolDG auf §§ 5 bis 9 GeoZG und damit u.a. auf die Vorgaben zum Mindestinhalt von Metadaten (§ 7 Abs. 2, Abs. 3 GeoZG). Für die öffentliche Bereitstellung nichtstaatlicher Geologiedaten enthalten die §§ 26 ff. GeolDG Sonderregelungen.

Die explizit geregelten und auch für Behörden geltenden ([34], S. 61) Datenübermittlungspflichten an eine zentrale für die Veröffentlichung zuständige Stelle sind eine Besonderheit des GeolDG. Die §§ 8 ff. GeolDG machen detaillierte Vorgaben dazu, welche Daten an die zuständige Behörde zu übermitteln sind und sehen hierfür auch Fristen vor. Diese Regelungen könnten als Anregung dienen, wenn es darum geht, die Übermittlung der für den Betrieb von umwelt.info erforderlichen Informationen/Metadaten zu Verfügung zu stellen – unabhängig davon, ob eine gesetzliche Regelung oder eher eine Übereinkunft in Staatsverträgen oder Verwaltungsabkommen angestrebt werden (vgl. Kap. 7.2.2).

### **7.1.5 E-Government Gesetz (Open Data: §12a EGovG)**

Der Anwendungsbereich der Datenbereitstellungspflicht nach § 12a EGovG wurde mit dem Zweiten Open Data Gesetz vom 16.07.2021 erweitert. Die Vorschrift bindet nunmehr alle Behörden des Bundes mit Ausnahme der Selbstverwaltungskörperschaften; außerdem sind auch Forschungsdaten erfasst. Das BMI wird in § 12a Abs. 12 EGovG ermächtigt, im Einvernehmen mit den übrigen Ministerien die Einzelheiten zum Bereitstellungsprozess durch Rechtsverordnung zu regeln; dies ist bislang aber noch nicht geschehen.

Die Regelung ist für die Umsetzung von umwelt.info insofern mittelbar relevant, als nun weitere Umweltdaten online zu stellen und damit verfügbar sind und über das Portal verlinkt werden können. Sollte eine Rechtsverordnung zur Datenbereitstellung ergehen, könnten sich hieraus Vorgaben (z. B. zur Bereitstellung von Metadaten) ergeben, die auch für umwelt.info relevant sein können.

### **7.1.6 Datennutzungsgesetz (DNG)**

Mit der soeben dargestellten Erweiterung des § 12a EGovG wurde auch das neue DNG erlassen, welches das bis dahin geltende Informationsweiterverwendungsgesetz (IWG) ablöste. Das DNG zielt (wie schon das IWG) auf eine im Grundsatz freie Weiterverwendung von Daten von öffentlichen Datenbereitstellern (§ 2 Abs. 2 DNG), sofern nicht Rechte Dritter (z. B. Urheberrechte) betroffen sind. Lizenzen müssen durch ein im Allgemeininteresse liegendes Ziel gerechtfertigt sein (§ 4 Abs. 3 DNG). Das Gesetz selbst gibt keinen Anspruch auf Zugang zu Daten, sondern erfasst nur Daten, für die entweder ein Zugangsanspruch besteht oder die faktisch öffentlich verfügbar sind.

Für die Umsetzung von umwelt.info kann das DNG etwa im Bereich redaktioneller Inhalte relevant sein (siehe Kap. 7.2.3). Das Gesetz gilt nicht für Umweltinformationen, soweit diese uneingeschränkt, kostenlos, maschinenlesbar und über eine API nutzbar sind (§ 2 Abs. 3 c) DNG). Unter diesen Bedingungen sind dann auch die Anforderungen des DNG – insbesondere die freie Weiterverwendung – automatisch gewahrt. Im Umkehrschluss zu dieser Ausnahme gilt das DNG für Umweltinformationen, welche die genannten Bedingungen nicht erfüllen.

### **7.1.7 Gegenwärtige Regulierungsbestrebungen auf europäischer Ebene**

Derzeit laufen auf EU-Ebene Gesetzgebungsprozesse, die (zumindest unter anderem) auf eine verbesserte Verfügbarkeit von Daten zielen; zu nennen sind hier insbesondere die Entwürfe für

einen Data Act sowie einen Data Governance Act (DGA). Beide Regularien beziehen sich nicht spezifisch auf Umweltinformationen, würden aber auch für diese gelten.

Der Data Act-Entwurf regelt den Zugang zu privat gehaltenen Daten. Konkret geht es um den Datenaustausch sowohl zwischen Unternehmen als auch von Unternehmen an die öffentliche Hand. Allerdings sollen Behörden nur in Ausnahmefällen Zugriffsrechte haben („Exceptional Need“), etwa in notstandsähnlichen Situationen („Public Emergency“) oder wenn eine Behörde eine konkrete gesetzliche Aufgabe ohne die Daten nicht wahrnehmen kann und diese auch nicht anderweitig (z. B. durch Kauf) beschaffen kann. Der Anwendungsbereich ist also sehr schmal.

Der DGA-Entwurf schlägt in Art. 3 ff. hingegen auch neue Regelungen für staatlich gehaltene Daten vor. Danach soll es staatlichen Stellen künftig erleichtert werden, auch geschützte Daten (z. B. aufgrund von Geschäftsgeheimnissen, Urheber- oder Datenschutzrechten), die nicht unter Open Data-Vorgaben fallen, zu transparenten Bedingungen verfügbar zu machen. Allerdings etabliert der DGA ausdrücklich keine Verpflichtung, sondern nur eine Möglichkeit, diese Daten zur Verfügung zu stellen, Art. 3 Abs. 3 DGA.

Die aktuellen Bestrebungen auf EU-Ebene können vor allem insofern (mittelbar) relevant werden, als sie möglicherweise neue Datenquellen für umwelt.info erschließen. Dies hängt von der Reichweite der letztlich verabschiedeten Verordnungstexte ab.

## **7.2 Rechtliche Anforderungen an die Umsetzung von umwelt.info im Überblick**

Im Folgenden werden auf Grundlage der Machbarkeitsstudie und des seitdem weiter fortgeschrittenen Umsetzungskonzepts die rechtlichen Anforderungen an den Betrieb von umwelt.info im Überblick dargestellt. Danach soll umwelt.info zunächst als Metadatensuchmaschine angelegt werden. Künftig sollen ggf. auch im Portal selbst Informationen hinterlegt werden. Außerdem sollen über das Portal redaktionelle Inhalte bereitgestellt werden.

Hinsichtlich der rechtlichen Anforderungen kann unterschieden werden zwischen den verlinkten Inhalten (Kap. 7.2.1), der Beschaffung, Haltung und Verarbeitung von Metadaten (Kap. 7.2.2), redaktionellen Inhalten sowie (ggf. künftig) selbst vorgehaltenen Daten (Kap. 7.2.3) und dem allgemeinen Portalbetrieb (Kap. 7.2.4).

Abschließend wird in Kap. 7.2.5 das Thema Verwaltungsvereinbarungen beleuchtet.

### **7.2.1 Keine rechtliche Verantwortlichkeit für verlinkte Inhalte**

Eine zentrale Weichenstellung der derzeitigen Konzeption besteht darin, dass das Portal als Metadatensuchmaschine angelegt wird, also vor allem über Hyperlinks auf externe Inhalte verweist. Diese Inhalte und die darin enthaltenen Daten sind weiterhin bei anderen Behörden (oder ggf. Privaten) hinterlegt, sodass das UBA als portalbetreibende Stelle hierfür grundsätzlich keine rechtliche Verantwortung trifft.

Insbesondere ist die bloße Verlinkung von Inhalten weder eine urheberrechtlich relevante Handlung noch eine Verarbeitung personenbezogener Daten. Verstöße gegen Datenschutz- und Urheberrecht (oder z. B. gegen den Schutz von Geschäftsgeheimnissen) auf verlinkten Webseiten sind der portalbetreibenden Stelle daher nicht zuzurechnen. Von diesem Grundsatz sind Ausnahmen denkbar, beispielsweise wenn technische Vorkehrungen zum Zugriffsschutz umgangen werden oder klar erkennbar urheberrechtswidrige Inhalte verlinkt werden. Allerdings hat das UBA insoweit keine Nachforschungspflichten, sondern muss nur auf konkrete Beanstandungen reagieren. Damit darf das UBA jedenfalls für die allgemeine Öffentlichkeit

kostenfrei bereitgestellte Informationen verlinken; dies betrifft z. B. auf Grundlage von § 10 UIG, § 12a EGovG oder des GeoZG (bzw. entsprechender Landesregelungen) bereitgestellter Daten.

Die dargestellten Grundsätze gelten sowohl für Verlinkungen auf behördliche als auch auf private Daten und Datensammlungen. Sie gelten ferner für die Anzeige von Vorschaubildern; auch insoweit hat der Suchmaschinenbetreiber keine Nachforschungspflicht, sondern muss nur reagieren, wenn er auf Rechtsverletzungen (also die unberechtigte Veröffentlichung des Bildes auf der verlinkten Seite) hingewiesen wurde ([35], Rn. 57 ff.).

Auch sonstige Anforderungen, z. B. Herausgabepflichten nach dem UIG (sowie dem IFG), sind vom tatsächlichen „Verfügen“ über die Daten abhängig (§ 2 Abs. 4 Satz 1 UIG), was nicht der Fall ist, wenn lediglich Umweltinformationen verlinkt werden, die bei anderen Stellen vorliegen. Anders ist dies nur, wenn das UBA selbst einen rechtlichen Anspruch (z. B. aus Gesetz oder Vertrag) gegen die jeweilige Stelle auf Herausgabe der betreffenden Daten hat (vgl. § 2 Abs. 4 Satz 2 UIG). Weil die verlinkten Daten nicht beim UBA/bei umwelt.info gespeichert sind, gelten auch die Open Data-Anforderungen nicht (vgl. § 12a Abs. 1, Abs. 2 Nr. 1 EGovG); umwelt.info verlinkt lediglich auf (verpflichtende) Open Data-Angebote.

Eine von der bloßen Verlinkung zu trennende Frage ist, inwieweit eine Bewertung der Quellenqualität der verlinkten Daten möglich ist (z. B. durch ein Ampelsystem). Grundsätzlich gilt, dass auch kritische staatliche Informationen zu privaten Produkten, Tätigkeiten usw. zulässig sind, wenn sie im Rahmen der eigenen Zuständigkeit geäußert werden sowie sachlich gehalten und zutreffend sind [36]. Dies lässt sich wohl grundsätzlich auf eine Bewertung behördlicher Angebote (Datenportale) übertragen. Soweit die Aufgabe umwelt.info als Verbundportal zu betreiben, in den Zuständigkeitsbereich von BMUV/UBA fällt, dürfte also auch eine sachliche Bewertung der verlinkten Quellen zulässig sein. Wichtig ist in jedem Fall eine Erläuterung, wie eine Bewertung zustande kommt.

## **7.2.2 Anforderungen an die Beschaffung, Haltung und Verarbeitung von Metadaten**

Rechtlich verantwortlich ist das UBA hingegen für Metadaten, weil und soweit es diese selbst als portalbetreibende Stelle beschafft, hält oder verarbeitet. Die Haltung und Verarbeitung von Metadaten einschließlich ihrer Beschaffung über sog. Harvester und Crawler ist rechtlich allerdings weitgehend unproblematisch.

Inhaltlich gibt es für Metadaten nur vereinzelt gesetzliche Vorgaben. Relevant sind vor allem die Vorgaben in § 7 GeoZG für Geodaten, die bestimmte Metadaten für die datenbereitstellenden Behörden vorgeben, die dann von umwelt.info übernommen werden können.

Metadaten selbst sind in aller Regel nicht urheberrechtlich geschützt, weil es an der für einen Urheberrechtsschutz notwendigen kreativen Leistung, der sog. „Schöpfungshöhe“ fehlt. Dies gilt immer dann, wenn Metadaten – wie bei Geodaten, vgl. § 7 GeoZG – gesetzlich vorstrukturiert sind, aber auch allgemein für „Kernmetadaten“, also einzelne Schlagworte und Schlüsselwörter. Ausnahmen sind lediglich bei umfangreicheren Beschreibungen denkbar, die Raum für kreative Gestaltungsfreiheit lassen [37], was wohl eher eine theoretische Möglichkeit ist. Vor diesem Hintergrund ist ein „Abgreifen“ durch das Kopieren von Metadaten (Harvester) kein Eingriff in Urheberrechte.

Vor diesem Hintergrund ist auch die Veränderung bzw. Ergänzung von Metadaten zu betrachten. Wie dargestellt, gibt es für Metadaten in der Regel weder inhaltliche Vorgaben noch genießen sie urheberrechtlichen Schutz. Daher ist auch die Veränderung von Metadaten nur dann problematisch, wenn sie entweder gegen inhaltliche Vorgaben verstößt (z. B., indem Pflichtangaben nach § 7 GeoZG gelöscht würden) oder urheberrechtliche Vorschriften

missachtet (z. B. Löschung/Veränderung des Urhebers in den Metadaten (siehe [38])). Demgegenüber ist die bloße Ergänzung von Metadaten grundsätzlich unproblematisch. Ausnahmen sind denkbar, vor allem, wenn personenbezogene Informationen ergänzt werden, die der Datensatz und dessen Metadaten bislang nicht enthielten; derartige Daten sollten nicht hinzugefügt werden.

Datenschutzrechtlich können Metadaten relevant sein, wenn sie selbst personenbezogen sind oder Rückschlüsse auf personenbezogene Daten zulassen. Personenbezogen sind etwa die Namen der Autor\*innen von Studien (da umwelt.info allerdings ohnehin nur bereits veröffentlichte Werke verlinkt, ist ein solcher Namenshinweis in Metadaten regelmäßig unproblematisch). Personenbezogen können darüber hinaus auch andere Metadaten sein, beispielsweise grundstücksbezogene Daten wie Flurstücke oder Hausnummern. Je granularer derartige Daten sind, desto eher besteht ein Personenbezug. Soweit die Angabe eines geografischen Standorts in Metadaten gesetzlich vorgegeben ist (vgl. etwa § 7 Abs. 2 Nr. 3 GeoZG), ist eine Übertragung dieser Metadaten in umwelt.info grundsätzlich unproblematisch, da diese Daten ja ohnehin öffentlich bereitzustellen sind bzw. vor der Bereitstellung bereits auf entgegenstehende datenschutzrechtliche Belange geprüft werden mussten (vgl. § 12 Abs. 2 GeoZG). Werden Metadaten aus unstrukturierten Daten ausgelesen, kann eine nähere datenschutzrechtliche Prüfung unter Umständen sinnvoll sein, z. B., wenn es um geografische Angaben geht. Allerdings ist auch im Zusammenhang mit Metadaten das oben dargestellte „Suchmaschinenprivileg“ zu beachten, nach dem Suchmaschinen als bloße Vermittler nicht für die Inhalte haften, auf die sie verweisen. Metadaten beschreiben diese Datensätze und werden daher nur dann personenbezogene Daten enthalten, wenn auch der Datensatz selbst solche Daten enthält (z. B. ein Flurstück, eine Hausnummer o.Ä., Ansprechpersonen). Da für die Inhalte verlinkter Seiten ohne Kenntnis der Rechtswidrigkeit einer Veröffentlichung/Verarbeitung keine Haftung besteht (siehe auch Kap. 7.2.1), müssen auch die Metadaten nur bei entsprechenden Hinweisen überprüft werden.

Die bei umwelt.info vorhandenen Metadaten zu verlinkten Umweltinformationen stellen auch selbst Umweltinformationen im Sinne des UIG dar. Das bedeutet vor allem, dass sie auf Antrag herauszugeben sind (§ 3 Abs. 1 UIG), sofern nicht im Einzelfall entgegenstehende öffentliche oder private Belange überwiegen (§§ 8 und 9 UIG). Letzteres dürfte bei Metadaten, die lediglich der Auffindbarkeit von weiteren (Umwelt-)Informationen dienen, kaum einmal der Fall sein. Auch eine freiwillige proaktive Veröffentlichung (vgl. § 10 Abs. 1 UIG) von Metadaten ist daher – jedenfalls in aller Regel – unproblematisch.

Bei der Beschaffung von Metadaten stellt sich über die technische Umsetzung durch Harvester und Crawler hinaus die Frage, ob und wie Vereinbarung zur Erstellung und Weitergabe von Metadaten für umwelt.info getroffen werden sollten. Hierfür kommen grundsätzlich Bundesgesetze (vgl. Art. 91c GG), Staatsverträge und Verwaltungsabkommen in Betracht (siehe auch noch Kap. 7.2.5). Bundeseinheitliche Datenübermittlungspflichten für bestimmte Umweltdaten wurden zuletzt mit dem GeoIDG eingeführt (vgl. Kap. 7.1.4). Im Kontext der (technischen) Beschaffung von Metadaten ist weiterhin zu berücksichtigen, dass eine Überlastung der adressierten Webseite drohen kann, wenn Harvestern oder Crawlern zu häufig/intensiv auf sie zugreifen. Inwieweit eine solche Überlastung droht, hängt von der Zahl der Zugriffe und der betroffenen Webseite ab. Rechtlich problematisch wären vor allem vorsätzliche Überlastungen, um die es hier nicht geht. Außerhalb dessen ist wohl vor allem relevant, inwieweit eine Überlastung der Webseite typischerweise droht oder konkret



vorhersehbar ist<sup>9</sup>. Wichtig ist zudem, dass evtl. bestehende technische Schutzvorkehrungen (z. B. robots.txt-Dateien) nicht umgangen werden. Die Häufigkeit und Intensität, mit der Webseiten von umwelt.info durchsucht werden, lässt sich auch in Verwaltungsvereinbarungen festhalten.

Allgemein ist zudem eine regelmäßige Aktualisierung von Metadaten vorzunehmen, um ihre Richtigkeit über die Zeit sicherzustellen.

### **7.2.3 Verantwortlichkeit für redaktionelle Inhalte und selbst vorgehaltene Daten**

Das UBA ist rechtlich verantwortlich, soweit es auf umwelt.info eigene redaktionelle Inhalte bereitstellt. Dies können zum einen Texte, Videos, Bilder usw. sein, die unmittelbar auf der Webseite eingestellt werden, aber auch extern eingebundene Quellen (z. B. Twitter).

Zu beachten ist bei der Gestaltung redaktioneller Inhalte vor allem das Urheberrecht.

- ▶ Dies gilt allerdings nur eingeschränkt für externe („eingebettete“) Inhalte anderer Plattformen (z. B. Twitter). Denn hierbei handelt es sich um eine bloße Verlinkung, die nur dann urheberrechtlich relevant ist, wenn die Inhalte offensichtlich rechtswidrig sind oder Schutzvorkehrungen umgangen werden (Vgl. zum „Embedding“ und Schutzmaßnahmen zuletzt EuGH, Urt. v. 09.03.2021, Az. C 392/19, Rn. 33 ff. [39]; siehe insgesamt auch bereits in 7.2.1)
- ▶ Die Bereitstellung fremder Werke unmittelbar auf umwelt.info ist nur in den Grenzen des Urheberrechts zulässig. Dies ist insbesondere bei der Verwendung von Fotografien relevant, weil alle Fotografien – unabhängig von einer besonderen kreativen Leistung – zumindest als Lichtbilder geschützt sind (§ 72 UrhG). Dies bedeutet, dass vor einer Veröffentlichung die Zustimmung der Rechteinhaber\*innen einzuholen ist.
- ▶ Daten als solche sind hingegen nicht urheberrechtlich geschützt. Daher ist z. B. die Nutzung von Daten (etwa aus Statistiken) zur Erstellung von Diagrammen, die auf umwelt.info bereitgestellt werden, unproblematisch.

Umgekehrt können selbst erstellte redaktionelle Inhalte (Texte, Bilder, Videos usw.) ebenfalls urheberrechtlich geschützt sein. Insofern stellt sich die Frage, zu welchen Bedingungen das UBA eine Weiterverwendung durch Dritte zulassen möchte. Nach dem Datennutzungsgesetz (DNG) sind nutzungsbeschränkende Lizenzen rechtfertigungsbedürftig (siehe Kap. 7.1.6). Daher empfiehlt es sich grundsätzlich offene Lizenzen zu verwenden, die als Bedingung für die Weiterverwendung nur die Nennung der Urheberschaft vorsehen (z. B. Datenlizenz Deutschland [40] oder CC-BY [41]). Grundsätzlich ist die Nutzung der Datenlizenz Deutschland (DL-DE->BY-2.0 [40]) zu empfehlen, weil sie (anders als die CC-BY-Lizenz) für die deutsche Rechtslage konzipiert ist<sup>10</sup>. Verwendet das UBA bei der Erstellung redaktioneller Inhalte (oder bei der Bereitstellung eigener Daten) auch geschützte Inhalte bzw. Daten Dritter (kombiniert oder verschneidet also Datensätze), so müssen diesbezügliche Lizenzbedingungen beachtet werden. Dies kann ggf. dazu führen, dass der kombinierte Datensatz nicht unter einer offenen Lizenz bereitgestellt werden darf. Inwieweit Lizenzen kompatibel sind, ist im Einzelfall zu prüfen. Häufig ist zumindest eine Bereitstellung des Gesamtdatensatzes zu den strengeren Lizenzbedingungen möglich.

---

<sup>9</sup> Dies ergibt sich aus allgemeinen Haftungsgrundsätzen, gerichtliche Entscheidungen hierzu sind nicht bekannt.

<sup>10</sup> Näher zu Lizenzfragen das Rechtsgutachten zum Projekt „Datennutzungskonzept und Dateninfrastruktur für sozial- und naturwissenschaftliche Datenbestände des Umweltbundesamtes“, Forschungskennzahl 3719121070, unter 2.2.3 (Veröffentlichung in Vorbereitung).

Soweit umwelt.info künftig auch selbst Primärdaten bereitstellen sollte, wäre das UBA als portalbetreibende Stelle für diese Daten vollständig rechtlich verantwortlich, da die „Privilegien“ einer Linklösung (s.o., Kap. 7.2.1) nicht mehr gälten. Dies bedeutet, dass für bei umwelt.info selbst vorgehaltene Daten eine Verantwortlichkeit für die Einhaltung datenschutz- und urheberrechtlicher Anforderungen besteht. Ferner sind z. B. auch die Anforderungen des UIG, § 12a EGovG und des DNG zu beachten. Für die rechtlichen Anforderungen im Einzelnen wird auf die Ausführungen im Rechtsgutachten aus dem Projekt „Datennutzungskonzept und Dateninfrastruktur für sozial- und naturwissenschaftliche Datenbestände des Umweltbundesamtes“ verwiesen<sup>11</sup>. Etwaige Lizenzen für die weitere Nutzung von Daten, die von Dritten auf umwelt.info bereitgestellt werden sind zu beachten; die Lizenzbedingungen sind mit den Daten bereitzustellen, sodass Nutzende wissen, wie sie die jeweiligen Daten verwenden dürfen.

### **7.2.4 Verantwortlichkeit für den allgemeinen Portalbetrieb**

Als portalbetreibende Behörde muss das UBA die rechtlichen Anforderungen beachten, die mit dem Betrieb des Portals selbst zusammenhängen. Dies sind insbesondere datenschutzrechtliche Anforderungen, die sich aus der DSGVO, dem BDSG sowie dem neuen Telekommunikation-Telemedien-Datenschutz-Gesetz (TTDSG) ergeben. Datenschutzrechtliche Vorgaben sind sowohl für Nutzungsdaten zu beachten – also Daten, die bei der Nutzung anfallen, wie z. B. die IP-Adresse und der Zeitraum des Zugriffs – als auch für Bestandsdaten, die von registrierten Nutzenden erfasst werden (z. B. Name, Emailadresse).

Im Hinblick auf die Nutzungsdaten gilt etwa, dass das Setzen von Cookies und Analytic-Tools nur mit Einwilligung der Nutzenden möglich ist, wie nun durch § 25 TTSG eindeutig klargestellt ist (s.o., Kap. 7.1.3). Für Bestandsdaten, die im Zusammenhang mit der Registrierung erhoben werden, ist eine Verarbeitung der Daten zur Durchführung der vertraglichen Beziehung zulässig (Art. 6 Abs. 1 lit. b DSGVO); außerdem kann und sollte bei Erstellung des Nutzungskontos eine datenschutzrechtliche Einwilligung eingeholt werden (Art. 6 Abs. 1 lit. e DSGVO).

Neben dem Datenschutz greifen auch sonstige allgemeine Anforderungen an Online-Angebote, etwa zur Barrierefreiheit (BGG, BITV 2.0), z. B. die Bereitstellung von Grundinformationen in leichter Sprache.

Werden im Rahmen des Portalbetriebs Informationen über die Nutzung des Portals erhoben und z. B. statistisch aufbereitet, dürften diese keine Umweltinformationen darstellen (kein Fall des § 2 Abs. 3 UIG). Für die Rohdaten kann eine Pflicht zur Veröffentlichung nach § 12a EGovG bestehen; dies gilt allerdings nicht mehr für z. B. in Statistiken aufbereitete Daten. Insofern kann aber eine Herausgabepflicht auf Antrag nach dem Informationsfreiheitsgesetz (IFG) bestehen, da es sich um amtliche Informationen gemäß § 2 Nr. 1 IFG handelt.

Schließlich gibt es rechtlich keine Open-Source-Verpflichtung. Es ist allerdings ohne Weiteres zulässig und auch wünschenswert, entwickelte Komponenten offen bereitzustellen.

### **7.2.5 Verwaltungsvereinbarungen zur (Meta-)Datenbereitstellung**

Über umwelt.info sollen umweltbezogene Daten möglichst umfassend und vor allem – nämlich soweit sie nicht ausnahmsweise von umwelt.info selbst vorgehalten werden – durch eine Verlinkung bestehender Portale von (Landes-)Behörden verfügbar gemacht werden (s.o., Kap. 7.2.1 sowie 7.2.2). Dies setzt voraus, dass Daten von diesen Behörden überhaupt bereitgestellt werden, da sie nur in diesem Umfang verlinkt werden können. Insofern besteht das Problem,

---

<sup>11</sup> Forschungskennzahl 3719121070.



dass gegenwärtig nur ein kleiner Teil der verfügbaren Umweltinformationen online zur Verfügung gestellt werden muss und daher Lücken bestehen. Da der Erlass umfassender gesetzlicher Datenbereitstellungs- und -übermittlungspflichten zugunsten von umwelt.info wenig realistisch erscheint, kommen für die Schließung von Datenlücken vor allem Verwaltungsvereinbarungen in Betracht. Für Vereinbarungen zum Datenaustausch zwischen Bundes- und Landesbehörden bietet Art. 91c Abs. 1, Abs. 2 GG eine verfassungsrechtlich abgesicherte Grundlage ([42], S. 9). Vereinbarungen sind also zulässig, ihr Inhalt hängt letztlich von den Interessen der Beteiligten ab.

Soweit es bloß um die Verfügbarkeit und Verlinkung von Informationen geht, könnte eine Vereinbarung etwa auf die Veröffentlichung durch die Landesbehörde und das Einverständnis zur Bereitstellung über umwelt.info beschränkt sein; das UBA als portalbetreibende Behörde würde die Informationen dann – wie auch sonst – selbst „abgreifen“ und verlinken. Sinnvoll ist es, die bereitzustellenden Daten möglichst genau zu bestimmen. Auch zur Aktualität von Daten und Zeitnähe ihrer Veröffentlichung können Festlegungen getroffen werden, ebenso zum Abruf der Metadaten durch umwelt.info (z. B. Zeitpunkt und Häufigkeit des Zugriffs über Harvester/Crawler, dazu oben, 7.2.2). Es sollte vereinbart werden, dass keine geschützten Daten (Urheberrecht, Datenschutzrecht, Geschäftsgeheimnisse) veröffentlicht werden, um sich insoweit abzusichern. Anders als sonst bei einer Linklösung wäre die Veröffentlichung ja hier durch das UBA veranlasst, weshalb sich eine vertragliche Absicherung empfiehlt. Umgekehrt sollte das UBA erklären, dass es bei der Bereitstellung über umwelt.info etwaige Nutzungsbedingungen beachtet, insbesondere Nutzungsbeschränkungen nicht umgeht. Denkbar wäre ferner die Vereinbarung von Standards für Metadaten, soweit dies sinnvoll und eine entsprechende Bereitschaft der Länder gegeben ist.

Denkbar wäre grundsätzlich auch eine Übersendung von Daten zur Veröffentlichung unmittelbar auf umwelt.info. Dann trägt das UBA selbst die Verantwortung für die Einhaltung rechtlicher Vorgaben (Urheberrecht, Datenschutz usw., s.o., 7.2.3). In der Vereinbarung sollten die Landesbehörden jedenfalls erklären, dass sie die Rechte zur Bereitstellung auf umwelt.info übertragen und hierzu auch berechtigt sind. Auch eine Erklärung, dass andere Anforderungen (Datenschutz u.Ä.) der Veröffentlichung nicht entgegenstehen, ist grundsätzlich sinnvoll (sofern Landesbehörden hierzu bereit sind). Ansonsten kann auch hier – in verschiedener Detailtiefe – festgelegt werden, welche Daten wie häufig und in welcher Weise (z. B. Format) zu übermitteln sind.

## 7.3 Handlungsempfehlungen

Im Folgenden werden für die oben angesprochenen Handlungsfelder Empfehlungen bzw. Leitlinien formuliert. Konkret betrifft dies die Verlinkung von externen Datenbeständen (Kap. 7.3.1), die Bereitstellung von Daten und Inhalten unmittelbar auf umwelt.info (7.3.2), den allgemeinen Portalbetrieb (Kap. 7.3.3) sowie den Abschluss von Verwaltungsvereinbarungen zum Datenaustausch (Kap. 7.3.4).

### 7.3.1 Verlinkung von Portalen und Datensammlungen Dritter

Bei der Verlinkung von Portalen und Datensammlungen Dritter und der damit verbundenen Beschaffung (z. B. Harvesten) von Metadaten sind folgende Aspekte zu berücksichtigen:

- Das UBA muss verlinkte Portalinhalte nicht aktiv auf Datenschutz- oder Urheberrechtsverstöße untersuchen. Konkreten Beanstandungen muss allerdings nachgegangen werden. Hierfür kann eine Ansprechstelle für Beschwerden bzw. ein entsprechendes Kontaktformular eingerichtet werden. Hierauf sollte auf dem Portal (z. B. in

den Nutzungsbedingungen) hingewiesen werden. Eine rechtliche Überprüfung von Beschwerden ist sicherzustellen, im Zweifel sollte eine Verlinkung zunächst aufgehoben werden.

- ▶ Soll eine Bewertung der Quellenqualität erfolgen, so ist diese sachlich und nachvollziehbar zu gestalten. Bei Nutzung eines „Ampelsystems“ sollte etwa gewährleistet sein, dass für die Nutzenden leicht einsehbar ist, anhand welcher Bewertungskriterien die Bewertung erfolgt ist (z. B. durch einen Klick auf die Ampel).
- ▶ Das Sammeln („Harvesten“) von Metadaten frei zugänglicher Inhalte ist rechtlich grundsätzlich unproblematisch. Allenfalls das Datenschutzrecht kann in Ausnahmekonstellationen betroffen sein. In diesen Fällen wird auch der beschriebene Datensatz datenschutzrechtlich relevant sein; bei entsprechenden Beanstandungen (1. Spiegelstrich) sollten dann auch die Metadaten geprüft werden.
- ▶ Das Verändern von Metadaten kann unter Umständen rechtlich problematisch sein, etwa wenn Daten zur Urheberschaft oder Lizenzen gelöscht werden. Daher sollten Metadaten grundsätzlich nur ergänzt, nicht aber gelöscht/überschrieben werden. Personenbezogene Daten, die weder im Hauptdatensatz noch dessen Metadaten enthalten sind, sollten nicht hinzugefügt werden.
- ▶ Eine regelmäßige Aktualisierung von Metadaten bzw. die Überprüfung der Links auf externe Inhalte ist zu empfehlen.
- ▶ Auf Anforderung sind die von umwelt.info verwendeten Metadaten nach dem UIG herauszugeben, wenn nicht (was nur ausnahmsweise in Betracht kommt) öffentliche oder private Belange überwiegen. Eine proaktive Veröffentlichung ist unter diesen Umständen zulässig, aber nicht rechtlich geboten.

### 7.3.2 Redaktionelle Inhalte und Datenbereitstellung

Bei der Erstellung redaktioneller Inhalte und eigener Datenbereitstellung unmittelbar auf umwelt.info sind folgende Aspekte zu berücksichtigen:

- ▶ Für eigene redaktionelle Inhalte und unmittelbar auf umwelt.info bereitgestellte Daten haftet das UBA für die Einhaltung rechtlicher Vorgaben. Dies gilt insbesondere für die Einhaltung von Datenschutzrecht, Urheberrecht und Schutz von Geschäftsgeheimnissen. Daher muss das UBA für diesen Bereich entsprechende personelle Kapazitäten für die Prüfung dieser rechtlichen Anforderungen einplanen.
  - Insbesondere bei der Bereitstellung von Bildern ist Vorsicht geboten, weil diese einen weitgehenden urheberrechtlichen Schutz genießen. Hier ist stets die Einwilligung der Berechtigten einzuholen.
  - Bei der Bereitstellung von Datensätzen, die geografische Angaben (z. B. Flurstücke, Hausnummern) enthalten, ist darauf zu achten, dass diese ggf. datenschutzrechtlich geschützt sein können. Je granularer die Daten sind, desto eher kommt ein Schutz in Betracht.
- ▶ Werden von Dritten erhobene Daten unmittelbar auf umwelt.info bereitgestellt, empfiehlt es sich, in einer Vereinbarung mit dem oder der Dritten festzuhalten, dass er oder sie die für die Bereitstellung auf umwelt.info erforderlichen Berechtigungen überträgt und hierzu selbst berechtigt ist. Damit sichert sich das UBA gegen eine Haftung ab. Auch die

datenschutzrechtliche Einwilligung für die Verarbeitung personenbezogener Daten des oder der Dritten sollte hier eingeholt werden. Zu empfehlen ist auch die Zusicherung, dass die Daten keine personenbezogenen Daten Dritter und Betriebs- und Geschäftsgeheimnisse enthalten.

- Soweit externen Dritten ein eigenständiges Hochladen von Daten auf umwelt.info ermöglicht wird, können und sollten die genannten Punkte in einer Nutzungsvereinbarung festgehalten werden, der vor einem Hochladen der Dateien zugestimmt werden muss.
- Lizenzen und Nutzungsbedingungen Dritter zu den auf umwelt.info vom Dritten bereitgestellten Daten sind vom UBA zu beachten, auf ihre Geltung sollte unmittelbar bei den Daten hingewiesen werden.
- Jedenfalls selbst oder im Auftrag des UBA erhobene Daten sollten unter offenen Lizenzen bereitgestellt werden, die als Bedingung für die Weiterverwendung nur die Nennung der Urheberschaft vorsehen. Zu empfehlen ist die Datenlizenz Deutschland mit Namensnennung (DL-DE->BY-2.0).
  - Werden Daten aus verschiedenen Quellen kombiniert, so ist die Kompatibilität der jeweiligen Nutzungsbedingungen zu prüfen. Häufig wird eine Bereitstellung des „Gesamtprodukts“ unter den strengeren Nutzungsbedingungen der sicherste Weg sein.
- Für weitere Informationen/Hilfestellungen siehe das Rechtsgutachten im „DNK-Projekt“ (Forschungskennzahl 3719121070).

### **7.3.3 Allgemeiner Portalbetrieb**

Als Betreiber des Portals hat das UBA die betriebliche Verantwortlichkeit:

- Für den allgemeinen Portalbetrieb sind die Anforderungen zu beachten, die auch für andere vom UBA betriebene Portale gelten, bspw. das UVP-Portal. Insofern ist ggf. eine Orientierung an bestehenden Gestaltungen möglich.
  - Dies gilt beispielsweise für Impressum und Datenschutzerklärung
  - Die Einwilligung zur Setzung von Cookies ist zwingend erforderlich, soweit diese nicht für den Betrieb der Seite unbedingt erforderlich sind. Eine Ablehnung nicht notwendiger Cookies muss einfach möglich sein.
- Auf der Seite sollten allgemeine Hinweise und Nutzungsbedingungen bereitgestellt werden. Diese sollten z. B. klarstellen, dass die auf dem Portal verfügbaren Daten nur zu den jeweils festgelegten Lizenzen zulässig sind. Sinnvoll ist auch der Hinweis, dass keine Verantwortlichkeit für bloß verlinkte Datenquellen übernommen wird.
- Sofern eine Registrierung von Nutzenden ermöglicht wird, werden personenbezogene Daten erhoben, sodass eine datenschutzrechtliche Einwilligung (hinsichtlich der jeglicher Verwendung von personenbezogenen Daten) einzuholen ist. Wird Nutzenden zudem das Einstellen von Daten auf dem Portal gestattet, ist eine Zustimmung zu diesbezüglichen Nutzungsbedingungen einzuholen (z. B. Erklärung zu Berechtigungen, siehe dazu oben).

### **7.3.4 Verwaltungsvereinbarungen zur (Meta-)Datenbereitstellung**

Folgende Themen sollten bezüglich Verwaltungsvereinbarungen beachtet werden:

- ▶ **Verwaltungsvereinbarungen mit (Landes-)Behörden zur Datenbereitstellung über umwelt.info sind zulässig.** Sie können sowohl die Bereitstellung von Daten auf behördeneigenen Webseiten und deren Verlinkung auf umwelt.info als auch die Übermittlung von Daten zur unmittelbaren Bereitstellung über umwelt.info zum Gegenstand haben.
- ▶ **Mögliche Inhalte einer Vereinbarung zur Bereitstellung auf (landes-)behördeneigenen Webseiten zur Verlinkung über umwelt.info**
  - Eine nähere Bestimmung der Daten, die bereitgestellt werden sollen (z. B. anhand des Gegenstands oder der Rechtsgrundlage der Erhebung) sollte enthalten sein
  - Eine Verpflichtung der anderen Behörde, bei der Veröffentlichung auf ihrer Webseite Rechte Dritter (Datenschutz, Urheberrechte und Geschäftsgeheimnisse) zu achten, sollte enthalten sein
  - Das UBA muss etwaige Nutzungsbeschränkungen der verlinkten Daten bei der Bereitstellung über umwelt.info beachten
  - Der zeitliche Turnus der Bereitstellung von Daten auf der (landes-)behördeneigenen Webseite kann geregelt werden
  - Die Modalitäten des Metadatenabrufs durch umwelt.info kann ebenfalls geregelt werden, z. B. die Häufigkeit des Abrufs
  - Möglich sind auch weitere Vereinbarungen zur Bereitstellung der Daten (z. B. Formate, Metadaten o.Ä.)
- ▶ **Mögliche Inhalte einer Vereinbarung zur Übermittlung von Daten durch (Landes-)Behörden zur Bereitstellung auf umwelt.info**
  - Es sollte festgehalten werden, dass die übermittelnde Behörde die für die Bereitstellung auf umwelt.info erforderlichen Berechtigungen überträgt und hierzu selbst berechtigt ist. Damit sichert sich das UBA gegen eine Haftung ab. Zu empfehlen ist auch die Zusicherung, dass die Daten keine personenbezogenen Daten Dritter und Betriebs- und Geschäftsgeheimnisse enthalten.
  - Weiterhin sollte der Gegenstand der Übermittlungspflichten (d.h. welche Daten zu übermitteln sind) klar gefasst werden
  - Daneben können sich – je nach Konstellation in verschiedener Detailtiefe – Vereinbarungen zu Häufigkeit und Modalitäten (z. B. Formate) der Datenübertragung empfehlen
- ▶ **Stets sollten Regelungen zu Laufzeit und Kündigung der Vereinbarungen aufgenommen werden**

## 8 Quellenverzeichnis

- [1] G. Börner, M. Bluhm, T. Fechner, R. Illes, B. Lubahn, M. Ostkamp, S. Richter, M. Schromm, U. Voges, J. von Zadelhoff, H. Rudolf, L. Hantsche, E. M. Lütkemeyer, M. Zschiesche, M. Niebuhr und R. Nöske, „Umwelt- und Naturschutzinformationssystem UNIS-D - Machbarkeitsstudie,“ Umweltbundesamt, Dessau-Roßlau, 2021.
- [2] B. Wake, „INVEST in Good Stories, and SMART Tasks,“ 17 08 2003. [Online]. Available: <https://xp123.com/articles/invest-in-good-stories-and-smart-tasks/>. [Zugriff am 08 09 2021].
- [3] M. Cohn, „Split Stories Using SPIDR,“ [Online]. Available: <https://www.mountaingoatsoftware.com/exclusive/spidr-poster-download>. [Zugriff am 08 09 2021].
- [4] BMU, „Strukturwandel im Mitteldeutschen Revier - Aufbau des Umweltinformationszentrums in Merseburg,“ 26 08 2021. [Online]. Available: <https://www.bmu.de/pressemitteilung/strukturwandel-im-mitteldeutschen-revier-aufbau-des-umweltinformationszentrums-in-merseburg>. [Zugriff am 09 10 2021].
- [5] Dr. Gernot Starke, Dr. Peter Uruschka und Mitwirkende, „arc42, das Template zur Dokumentation von Software- und Systemarchitekturen.,“ [Online]. [Zugriff am 19.08.2021].
- [6] „Architekturrichtlinie für die IT des Bundes,“ Der Beauftragte der Bundesregierung für Informationstechnik, 2020. [Online]. Available: [https://www.cio.bund.de/Web/DE/Architekturen-und-Standards/Architekturrichtlinie-IT-Bund/architekturrichtlinie\\_it\\_bund\\_node.html](https://www.cio.bund.de/Web/DE/Architekturen-und-Standards/Architekturrichtlinie-IT-Bund/architekturrichtlinie_it_bund_node.html). [Zugriff am 08 09 2021].
- [7] „Europäisches Datenportal,“ [Online]. Available: <https://data.europa.eu>. [Zugriff am 01.09.2021].
- [8] bitkom, „www.bitkom.org,“ 2017. [Online]. Available: <https://www.bitkom.org/Bitkom/Publikationen/Entscheidungsunterstuetzung-mit-Kuenstlicher-Intelligenz.html>. [Zugriff am 01 09 2022].
- [9] Microsoft, „new.microsoft.com,“ Microsoft, 4 März 2020. [Online]. Available: <https://news.microsoft.com/de-at/microsoft-erklart-was-ist-kunstliche-intelligenz-definition-funktionen-von-ki/>. [Zugriff am 1 September 2022].
- [10] Europäisches Parlament, „www.europarl.europa.eu,“ 14 September 2020. [Online]. Available: <https://www.europarl.europa.eu/news/de/headlines/society/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt>. [Zugriff am 1 September 2022].
- [11] S. Raschka und V. Mirjalili, Machine Learning mit Python und Keras, TensorFlow 2 und SciKit-learn: Das umfassende Praxis-Handbuch für Data science, Deep Learning und Predictive Analytics, MITP-Verlags GmbH & Co. KG, 2021.
- [12] F. Chollet, Deep Learning with Python, Manning, 2017.
- [13] GyanSetu!, 25 Februar 2021. [Online]. Available: <https://www.gyansetu.in/what-is-natural-language-processing/>. [Zugriff am 2 September 2022].
- [14] Y. Goldberg, „A Primer on Neural Network Models for Natural Language Processing,“ *Journal of Artificial Intelligence Research*, Nr. 57, pp. 354-420, 2016.

- [15] V. S. Prem Melville, „Recommender Systems,“ in *Encyclopedia of Machine Learning*, Berlin, Springer-Verlag, 2010, pp. 1-9.
- [16] „Beispiel für SDW,“ [Online]. Available: <https://geo4web-testbed.github.io/topic4/#h.qah4wfgawhhl>. [Zugriff am 07.2021].
- [17] „w3.org / semanticweb,“ [Online]. Available: <https://www.w3.org/standards/semanticweb/data>. [Zugriff am 07.2021].
- [18] „Linked Open Data,“ [Online]. Available: [https://www.w3.org/egov/wiki/Linked\\_Open\\_Data](https://www.w3.org/egov/wiki/Linked_Open_Data). [Zugriff am 07.2021].
- [19] Umweltbundesamt, „Umweltbundesamt, Fachbibliothek Umwelt,“ 12 10 2022. [Online]. Available: [https://doku.uba.de/aDISWeb/app?service=direct/0/Home/\\$DirectLink&sp=SOPAC](https://doku.uba.de/aDISWeb/app?service=direct/0/Home/$DirectLink&sp=SOPAC). [Zugriff am 12 10 2022].
- [20] B. Ojokoh, „A Review of Question Answering Systems,“ *Journal of Web Engineering*, Bd. 17, Nr. 8, pp. 717-758, 2019.
- [21] A. M. Pundge, K. S.A. und C. N. Mahender, „Question Answering System, Approaches and,“ *International Journal of Computer Applications*, Bd. 141, Nr. 3, pp. 34-39, 2016.
- [22] J. Devlin, M.-W. Chang, K. Lee und K. Toutanova, „BERT: Pre-training of Deep Bidirectional Transformers for,“ *The North American Chapter of the Association for Computational Linguistics*, p. 4171–4186, 2019.
- [23] A. Bonde, „www.mycustomer.com,“ 24 August 2021. [Online]. Available: <https://www.mycustomer.com/marketing/data/what-is-ai-enabled-consumer-intelligence-aici-and-why-should-you-care>. [Zugriff am 2 September 2022].
- [24] R. C. Martin, *Clean Agile: Back to Basics*, Pearson, 2019.
- [25] J. Appelo, *Management 3.0: Leading Agile Developers, Developing Agile Leaders*, Boston: Pearson Education, Inc., 2011.
- [26] S. Ockerman und S. Reindl, *Mastering Professional Scrum: A Practitioner’s Guide to Overcoming Challenges and Maximizing the Benefits of Agility*, Addison-Wesley Professional, 2019.
- [27] M. Cohn, *Succeeding with Agile: Software Development Using Scrum*, Upper Saddle River, NJ: Addison-Wesley, 2010.
- [28] „Agile Manifesto,“ 2001. [Online]. Available: <https://agilemanifesto.org/iso/de/manifesto.html>. [Zugriff am 08 09 2021].
- [29] D. F. Termer und T. Zehler, *Agilität in Organisationen - Ein Leitfaden für die Praxis*, Bitkom Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., 2020.
- [30] E. Derby und D. Larsen, *Agile Retrospectives: Making Good Teams Great*, Raleigh, NC: Pragmatic Bookshelf, 2006.
- [31] „The 2020 Scrum Guide,“ 2020. [Online]. Available: <https://scrumguides.org/scrum-guide.html>. [Zugriff am 08 09 2021].
- [32] M. Burrows, *Kanban: Verstehen, einführen und anwenden*, Heidelberg: dpunkt.verlag GmbH, 2015.

- [33] R. Jeffries, „What is Extreme Programming?“, 16 03 2011. [Online]. Available: <https://ronjeffries.com/xprog/what-is-extreme-programming/>. [Zugriff am 26 10 2021].
- [34] M. Cohn, User Stories Applied: For Agile Software Development, Addison-Wesley Professional, 2004.
- [35] O. Wolf, „Agile Softwareentwicklung: Vertragsgestaltung und Praxistipps“, 23 05 2021. [Online]. Available: <https://www.ra-plutte.de/agile-softwareentwicklung-vertragsgestaltung-praxistipps/>. [Zugriff am 05 10 2021].
- [36] R. Jeffries, „Essential XP: Card, Conversation, Confirmation“, 30 08 2001. [Online]. Available: <https://ronjeffries.com/xprog/articles/expcardconversationconfirmation/>. [Zugriff am 08 09 2021].
- [37] G. Starke, Dr. und P. Uruschka, Dr., „arc42, das Template zur Dokumentation von Software- und Systemarchitekturen.“, 2017. [Online]. Available: <https://arc42.org/download>. [Zugriff am 19 08 2021].
- [38] „Metadata Quality Assessment Methodologie des European Data Portals“, [Online]. Available: <https://data.europa.eu/mqa/methodology?locale=de>. [Zugriff am 22.09.2021].
- [39] „FAIR Guiding Principles for scientific data management and stewardship“, [Online]. Available: <https://www.go-fair.org/fair-principles/>. [Zugriff am 22.09.2021].
- [40] „Data Quality Vocabular, W3C“, [Online]. Available: <https://www.w3.org/TR/vocab-dqv/>. [Zugriff am 22.09.2021].
- [41] „Kanban University Materials“, Kanban University, [Online]. Available: <https://resources.kanban.university/materials/>. [Zugriff am 26 10 2021].
- [42] R. Pichler, S. Roock und A. Beppe, Agile Entwicklungspraktiken mit Scrum, Heidelberg: dpunkt-Verlag, 2011.
- [43] K. Beck und C. Andres, Extreme Programming Explained: Embrace Change, Addison-Wesley Professional, 2004.
- [44] P. Resnick und H. R. Varian, „Recommender Systems“, *Communications of the ACM*, Bd. 40, Nr. 3, pp. 56-58, 1997.
- [45] F. Ricci, L. Rokach und B. Shapira, Recommender Systems Handbook, 2010.
- [46] J. Kaarma, „medium.com“, 21 Februar 2018. [Online]. Available: <https://medium.com/e-residency-blog/how-we-use-ai-to-help-users-get-answers-instantly-and-increase-customer-satisfaction-95d1a06caad>. [Zugriff am 2 September 2022].
- [47] T. Grainger, D. Turnbull und M. Irwin, AI Powered Search, Manning, im Druck.
- [48] K. Brush und J. Scardina, „www.techtarget.com“, [Online]. Available: <https://www.techtarget.com/searchcustomerexperience/definition/chatbot>. [Zugriff am 2 September 2022].
- [49] G. Caldarini, S. Jaf und K. McGarry, „A Literature Survey of Recent Advances in Chatbots“, *Information*, Bd. 13, Nr. 41, pp. 1-22, 2022.
- [50] onlim.com, „onlim.com“, 15 März 2022. [Online]. Available: <https://onlim.com/de/die-geschichte-und-entwicklung-von-chatbots/>. [Zugriff am 2 September 2022].



- [51] Engati Team, „[www.engati.com](http://www.engati.com),“ 20 August 2022. [Online]. Available: <https://www.engati.com/blog/what-are-chatbots>. [Zugriff am 2 September 2022].
- [52] „DCAT-AP-JRC,“ [Online]. Available: <https://ec-jrc.github.io/dcat-ap-jrc/>. [Zugriff am 05.10.2021].
- [53] „ATOM / RFC 4287,“ [Online]. Available: <https://datatracker.ietf.org/doc/html/rfc4287>. [Zugriff am 21.09.2021].
- [54] „RSS Spezifikation,“ [Online]. Available: <https://www.rssboard.org/rss-specification>. [Zugriff am 21.09.2021].
- [55] „OGC CSW AP ISO,“ [Online]. Available: [https://portal.ogc.org/files/?artifact\\_id=77855](https://portal.ogc.org/files/?artifact_id=77855). [Zugriff am 01.09.2021].
- [56] „INSPIRE,“ [Online]. Available: [https://inspire.ec.europa.eu/file/1551/download?token=Ws4F8\\_aS](https://inspire.ec.europa.eu/file/1551/download?token=Ws4F8_aS). [Zugriff am 01.09.2021].
- [57] „OGC API Records,“ [Online]. Available: <https://ogcapi.ogc.org/records/>. [Zugriff am 01.09.2021].
- [58] „dsgvo-gesetz.de,“ [Online]. Available: <https://dsgvo-gesetz.de/>. [Zugriff am 01.09.2021].
- [59] EEB, „Still too hard to access environmental information in the EU – EEB report,“ EEB, 02 12 2019. [Online]. Available: <https://eeb.org/still-too-hard-to-access-environmental-information-in-the-eu-eeb-report/>. [Zugriff am 27 10 2022].

## **A Anhang A: Systemarchitektur**

Die Beschreibung der Systemarchitektur der Ausbaustufen 1 und 2 von umwelt.info basiert auf dem Arc42 Architektur Template. Die Themen Technische Randbedingungen und Qualitätsziele, Kontextabgrenzung und Lösungsstrategie sind in Kap. 3 enthalten. Der Anhang A dient der detaillierten Beschreibung der Systemarchitektur für das Entwicklungsteam und enthält die Bausteinsicht mit der Zerlegung des Systems in Komponenten, die Laufzeitsicht, die Verteilungssicht sowie querschnittliche Konzepte und Entwurfsentscheidungen.

## **B Anhang B: Agiles Vorgehensmodell**

Im Umsetzungskonzept wurde ein Modell für das Vorgehen bei der Realisierung von umwelt.info entworfen und ein Organisationskonzept für Entwicklung und den späteren Betrieb erstellt. Das Vorgehensmodell ist in Kap. 5 zusammenfassend dargestellt und wird im Anhang B für die Nutzung durch das Entwicklungs- und Betriebsteam ausführlich beschrieben.

## C Anhang C: Glossar

### **Artefakt**

Ein Artefakt ist ein Ergebnis, das während des Entwicklungsprozesses erarbeitet wird. Die drei wesentlichen Artefakte in Scrum-basierten Prozessen sind das Product- und das Sprintbacklog sowie das Inkrement, also der aktuelle Entwicklungsstand des zu entwickelnden Produkts.

### **Aufgabe**

Aufgaben werden während der Sprintplanung aus dem Sprintziel abgeleitet und den Developern namentlich zugewiesen. Aufgaben werden in der Regel als Items in ein Sprint-Backlog eingetragen.

### **Auftraggeber**

Gibt das Projekt beim Auftragnehmer in Auftrag, kontrolliert und steuert das Projekt inhaltlich. Der Auftraggeber muss die Rolle des Chief Product Owners besetzen und stellt auch ein oder mehrere Teams, die selbst inhaltlich arbeiten.

### **Auftragnehmer**

Führt im Wesentlichen die Entwicklungsarbeiten durch, stellt ein\*e Projektleiter\*in als zentrale Kontaktstelle und organisiert ein oder mehrere Teams. Für jedes Team gibt es eine Vertretung, die den Chief Product Owner unterstützt.

### **Backlog Item**

Eintrag im Product Backlog oder in einem Sprint Backlog.

### **Backlog Refinement**

Pflege eines Backlogs als Vorbereitung künftiger Sprints.

### **Chief Product Owner**

Dem Chief Product Owner obliegt die inhaltliche Leitung des Gesamtprojekts. Gemeinsam mit den Product Owners der verschiedenen Teams pflegt er\*sie das Product Backlog und steuert damit die Reihenfolge, in der die Aufgaben abgearbeitet werden. In fachliche Fragen bzgl. der Priorisierung von Aufgaben oder der fachlichen Spezifikation ist er\*sie die höchste Entscheidungsebene.

### **CI/CD**

Dies sind die kombinierten Praktiken aus Continuous Integration und Continuous Delivery oder Deployment.

### **Cluster**

Mit einem Cluster werden mehrerer Objekte mit ähnlichen Eigenschaften zusammengefasst. Im Architekturdokument wird der Begriff Cluster in Zusammenhang mit den User Stories und der Infrastruktur genutzt. Dabei bezieht sich der Begriff bei den User Stories auf die einzelnen User Stories mit gleichem Inhalt, welche zu einem Cluster zusammengefasst werden. In der Infrastruktur werden mit dem Begriff Cluster mehrere Server-Maschinen in einem Kubernetes Cluster zusammen orchestriert.

### **Continuous Deployment**

Mit Continuous Delivery oder Deployment werden in der Regel Praktiken oder Prozesse der Softwareauslieferung zusammengefasst (siehe auch CI/CD, DevOps).

### **Continuous Integration**

Die kontinuierliche Integration beschreibt in der Softwareentwicklung das frühe und ständige Zusammenführen von Code oder Komponenten zu einem lauffähigen System, um die Qualität der Software sicherzustellen (siehe auch CI/CD, DevOps).

### **Common Weakness Enumeration**

Common Weakness Enumeration ist eine gemeinschaftlich entwickelte Liste von Software- und Hardware-Schwachstellen, die durch das U.S. Ministerium für Innere Sicherheit gesponsort wird (siehe auch Dependency Management / Check und <https://cwe.mitre.org>).

### **Definition of Done**

Kriterien, die für einen Backlogeintrag formuliert werden und mit Hilfe derer entschieden wird, ob ein Eintrag als fertig bearbeitet gilt.

### **Design Notes**

Design Notes sind eine textliche Beschreibung von Gestaltungsentwürfen. Die Beschreibung hilft bestimmte Strukturen, Abläufe und Aktionen genauer zu erläutern, welches anhand von Bildern allein nicht möglich ist.

### **DevOps**

Bessere Softwarequalität und kurze Auslieferungszeiten sollen durch enge Zusammenarbeit von Softwareentwicklung (Dev) und IT-Betrieb (Ops) erreicht werden (siehe CI/CD).

### **Epic**

Mit einem Epic werden die Anforderungen einer neuen Software auf einer übergeordneten Abstraktionsebene beschrieben. Es wurden aus den bereits ermittelten User Stories der Machbarkeitsstudie die Anforderungen abgeleitet und diese in Epics als Lösungsansatz beschrieben. Dabei beziehen sich die Teamziele in der Regel auf ein Epic.

### **Ereignis**

Geplante Veranstaltung mit definiertem Kreis von Teilnehmenden und festgelegtem Zweck. Ereignisse dienen i.d.R. der Bearbeitung oder Erstellung bestimmter Artefakte. Welche Ereignisse vorgesehen sind, ist in der Beschreibung des Vorgehensmodells festgelegt.

### **Gesamtteam**

Alle Personen, die mit der Realisierung oder dem Betrieb von umwelt.info direkt betraut sind. Teams sind Teil des Gesamtteams.

### **Gesamtteam-Ziel**

Ziel eines Sprints für das Gesamtteam. Das Gesamtteam-Ziel wird vom Chief Product Owner gemeinsam mit den Product Owners aus einem Eintrag der Roadmap abgeleitet und formuliert. Aus dem Gesamtteam-Ziel leiten sich die Sprintziele der beteiligten Teams ab.

### **Integrated Development Environment**

Eine integrierte Entwicklungsumgebung wird zur Softwareentwicklung eingesetzt und zeichnet sich durch die Integration mehrerer Werkzeuge und Programme in einer Umgebung aus, die zur effizienten Programmierung benötigt werden.

### **Inkrement**

Entwicklungsfortschritt, der durch die Arbeit im Rahmen eines Sprints erreicht wird. Besteht in der Regel aus einer neuen lauffähigen und demonstrierbaren Version des Produkts und definiert sich aus dem funktionalen Unterschied zur vorherigen Version.

### **Linked Data**

Unter Linked Data wird eine erweiterte Verknüpfung von Inhalten im Web verstanden, die auf Standardwebtechnologien aufbaut und um Maschinenlesbarkeit erweitert. Insbesondere wird die semantische Bedeutung von Verknüpfungen maschinenlesbar. Damit werden verbesserte Abfragen durch Computerprogramme ermöglicht mit dem Ziel, das Internet zu einer globalen Datenbank zu erweitern (nach [16]).

### **Minimum Viable Product**

Erste, minimal funktionsfähige Version eines zu erstellenden Produkts. Dieses dient als Grundlage für die weitere Entwicklung und wird kontinuierlich ausgebaut und verbessert.

### **Nationales Zentrum für Umwelt- und Naturschutzinformationen**

Zuständige Stelle für die inhaltliche Pflege und die Weiterentwicklung von umwelt.info.

### **Organisationseinheit**

z. B. Auftraggeber, Auftragnehmer, Sub-Auftragnehmer etc.

### **Personas**

Personas sind fiktive konzipierte Charaktere, welche verschiedene Nutzungstypen des Produktes repräsentieren. Personas helfen dabei, die Bedürfnisse, Erfahrungen, Verhaltensweisen und Ziele der Nutzenden zu verstehen und sich in diese hineinzuversetzen. Dadurch wird es den Projektbeteiligten ermöglicht, die Anforderungen der Nutzenden stets im Blick zu behalten und die Ziele danach auszurichten.

### **Protokoll**

Protokolle dokumentieren getroffene Entscheidungen.

### **Product Backlog**

Liste mit allen Arbeitspaketen, die an die einzelnen Teams übergeben werden, um das Projektziel zu realisieren. Zentrales Steuer- und Kontrollinstrument des Chief Product Owner.

### **Product-Backlog Item**

Einzelner Eintrag im Product Backlog

### **Product Owner**

Verantwortlich für die Pflege der Sprint-Backlogs eines Teams. Schnittstelle zum Chief Product Owner und damit zum Product Backlog.

### **Projektziel**

Ziel auf Gesamtteam-Ebene. Aus einem Projektziel können ein oder mehrere Sprintziele abgeleitet werden. Das Projektziel ist so detailliert beschrieben, wie es nötig und möglich ist und kann im Verlauf der Entwicklung durchaus angepasst werden, um auf veränderte Anforderungen oder Rahmenbedingungen zu reagieren. Das Projektziel wird aus der Vision abgeleitet. Dabei handelt es sich um eine komprimierte Formulierung des Kerns des Projektziels, die sich nicht ändern sollte.

### **Pull Request**

Mit Hilfe von Pull Requests können Entwickelnde Änderungswünsche an Code eines Versionsverwaltungssystems (Git o.ä.) erstellen und diskutieren, welches zu einem anderen Produkt oder Team gehört. So können über Kollaboration auf Quellcode Ebene über Grenzen hinweg entstehen, Wissen verteilt oder Ideen in Wert gesetzt werden.

### **Roadmap**

Übersicht des Projektablaufs mit möglichen Meilensteinen. Aus den Einträgen der Roadmap leiten sich die Gesamtteam-Ziele ab.

### **Scrum Master**

Ein Scrum Master wird oft als Agile Coach innerhalb des Scrum Frameworks angesehen. In der Regel arbeiten Scrum Master als dienende Führung für Teams und die Organisation und sind dabei ergebnisverantwortlich für die Effektivität des Scrum Teams, indem sie helfen, die Theorie von Scrum zu verstehen und diese in der Praxis umzusetzen.

### **Software Craftsmanship**

Mit diesem Begriff soll die handwerkliche Seite der Softwareentwicklung beleuchtet werden und weitere Aspekte, wie den Berufsethos beschrieben werden, aus dem sich gewisse Standards und Ansprüche an das Selbstverständnis der Softwareentwickler\*innen ergeben.

### **Sprint**

Bezeichnet eine Entwicklungsphase des Gesamt- oder eines Teilteams. Ein Sprint ist zeitlich begrenzt, mit einem eigenen (Zeit-)Budget versehen und hat ein formuliertes Ziel. Am Ende des Sprints soll das zu entwickelnde Produkt dem Projektziel signifikant stärker entsprechen als vorher. Die Veränderung des Produkts während des Sprints ist das Inkrement.

### **Sprint Backlog**

Liste der Aufgaben, die ein Team in einem Sprint abarbeiten soll.

### **Team**

Gruppe von Entwickler\*innen und Product Owner, die gemeinsam an der Umsetzung eines Teilbereichs des Gesamtprojektes arbeitet. Ein Team arbeitet in Sprints und auf der Basis eines eigenen Sprint Backlogs.

### **Team-Ziel**

Leitet sich aus Gesamtteam-Ziel ab und bezieht sich in der Regel auf ein Thema und ein Epic.

### **User Centered Design**

Bei der nutzerorientierten Gestaltung stehen die Anforderungen der (potenziell) Nutzenden einer Softwareanwendung im Vordergrund. Die Ziele und durchzuführenden Tätigkeiten der Nutzenden werden vor der Entwicklung mit unterschiedlichen Methoden detailliert erhoben. Dieser Prozess ist ein iteratives Vorgehen, der mehrere Phasen durchläuft und die Nutzungserfahrung (User Experience) verbessern soll.

### **umwelt.info**

Portal des Nationalen Zentrums für Umwelt- und Naturschutzinformationen für den Zugriff auf Umweltinformationen und -daten aus Deutschland. System, für dessen Realisierung in diesem Projekt die konzeptionelle Vorarbeit geleistet wird. Frühere Bezeichnung (Arbeitstitel) war Umwelt- und Naturschutz-Informationssystem – Deutschland (UNIS-D).



### **User Journey**

Eine User Journey beschreibt relevante Abläufe und damit verbundene Erfahrungen von Nutzenden bei der Interaktion mit einem Produkt. Idealerweise wird der ganzheitliche Blick vom Beginn bis zur Zielerreichung erfasst.

### **User Story**

User Stories sind potenzielle Anforderungen der Nutzenden in Form von kurzen, spezifischen und zielorientierten Aussagen mit der Struktur: "Als <Rolle> möchte ich <Ziel/Wunsch>, um <Nutzen>". Im UX-Kontext werden sie dazu eingesetzt, um eine vielfältige Sammlung unterschiedlicher Anforderungen der Nutzenden zu sammeln, die im weiteren Verlauf geclustert werden, um die relevantesten Anforderungen auszuarbeiten.

### **UX (User Experience)**

Die Nutzungserfahrung umfasst nach DIN EN ISO 9241 die Wahrnehmungen und Reaktionen einer Person, die sich aus der Verwendung und/oder der erwarteten Verwendung eines Produkts, eines Systems oder einer Dienstleistung ergeben (siehe auch [34]). UX kann Emotionen, Überzeugungen, Präferenzen, Wahrnehmungen, physischen und psychologischen Reaktionen, Verhaltensweisen und Leistungen der Nutzenden, die vor, während und nach der Nutzung auftreten beinhalten.

### **Vision**

Die Vision ist eine stark komprimierte Beschreibung des zu erstellenden Systems. Die Formulierung einer Vision enthält keine Details, sondern den funktionellen Kern des Systems. Die Vision sollte bis zur Fertigstellung des Systems nicht geändert werden. Bei Änderungen des Projektziels sollten keine Widersprüche zur Vision entstehen.