

Institut für Produktion und Industrielles Informationsmanagement

Universität Duisburg-Essen, Campus Essen
Fakultät für Wirtschaftswissenschaften
Universitätsstraße 9, 45141 Essen
Tel.: +49 (0) 201 18 34007

Arbeitsbericht Nr. 48

zugleich

KI-LiveS-Projektbericht Nr. 2

Überblick über potenzielle Quellen für Test- und Evaluierungsdaten eines KI-Labors im Rahmen des KI-LiveS-Projekts

Schagen, J. P. • Zelewski, S. • Haselhoff, T. • Schmitz, S. • Heeb, T.



Verbundprojekt KI-LiveS: KI-Labor für verteilte und eingebettete Systeme
Förderkennzeichen: 01IS19068

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

E-Mail: stephan.zelewski@pim.uni-due.de

Internet: <https://www.pim.wiwi.uni-due.de/team/stephan-zelewski/>

ISSN 1614-0842

Essen 2021

Alle Rechte vorbehalten.

Zusammenfassung

Das BMBF-Forschungsprojekt „KI-LiveS“ (KI-Labor für verteilte und eingebettete Systeme) verfolgt primär das Transferziel („Third Mission“), Erkenntnisse aus der universitären Erforschung Künstlicher Intelligenz (KI) besser in der gewerblichen Wirtschaft zu verankern, um dort Entwicklungen von innovativen Produkten, insbesondere Dienstleistungen anzuregen, die den Wirtschaftsstandort Deutschland nachhaltig stärken. In diesem Kontext befasst sich der vorliegende Projektbericht Nr. 2 des KI-LiveS-Projekts mit einem Überblick über potenzielle Datenquellen, die für Test- und Evaluierungszwecke im Hinblick auf die Entwicklung von KI-Applikationen und für den Aufbau eines „Data Repository“ für ein KI-Labor genutzt werden können.

Abstract

The BMBF research project ‘KI-LiveS’ (AI laboratory for distributed and embedded systems) pursues primarily the third-mission-based aim of a more effective implementation of the university research of Artificial Intelligence (AI) into trade and industry in order to stimulate the development of innovative products, especially services, which strengthen the business location Germany sustainably. In this context, this project report no. 2 of the KI-LiveS project deals with an overview of potential data sources, which can be used for testing and evaluation purposes of AI applications and for building a data repository for an AI lab.

Danksagung

Dieser Projektbericht entstand durch die Kooperation zahlreicher Personen, die am KI-LiveS-Projekt mitwirken. Dazu zählen neben den Verfassern des Projektberichts vor allem drei Personenkreise. Erstens haben verschiedene Mitglieder des Projektkonstortiums des KI-LiveS-Projekts durch ihr großes Engagement hinsichtlich der Datenquellenbereitstellung und -beschreibung maßgeblich zu diesem Projektbericht beigetragen. Vor allem gilt Herrn Univ.-Prof. Dr. Schiele ein großer Dank für seine umfassenden Auskünfte während des Entstehens dieses Projektberichts. Zweitens haben studentische Mitarbeiter des Instituts für Produktion und Industrielles Informationsmanagement substantielle redaktionelle Zuarbeiten zu diesem Projektbericht geleistet. Eine besondere Hervorhebung verdienen Frau Sabah Allam, Frau Svenja Fink, Frau Leonie Weber und Herr Manuel Bros. Sie haben die Verfasser dieses Projektberichts stets hervorragend unterstützt. Drittens gilt ein besonderer Dank den Praxispartnern des KI-LiveS-Projekts, die sich an der Erhebung von potenziellen Datenquellen aktiv beteiligt haben. Sie werden in diesem Projektbericht an späterer Stelle explizit vorgestellt.

Darüber hinaus fühlen sich die Mitglieder des KI-LiveS-Projektkonsortiums („Universitätspartner“) dem BMBF als Förderer des Drittmittel-Verbundprojekts sowie dem Deutschen Zentrum für Luft- und Raumfahrt e. V. (DLR) als zuständigem Projektträger für die großzügige finanzielle Projektförderung bzw. für die professionelle Projektbegleitung zu großem Dank verbunden.

Inhaltsverzeichnis

	<u>Seite</u>
1 Bedeutung von Datenquellen für das KI-LiveS-Projekt.....	1
2 Erhebung potenzieller Datenquellen	3
2.1 Identifizierung potenzieller Datenquellen.....	3
2.2 Vorstellung der Erhebungsmethode	5
2.3 Identifizierung von Stakeholdern für die Erhebung von Datenquellen.....	7
2.4 Durchführung der Datenquellenerhebung.....	8
3 Auswertung der Datenquellenerhebung.....	9
3.1 Übersicht über potenzielle Datenquellen	9
3.2 Beschreibung der potenziellen Datenquellen.....	10
3.3 Übersicht über weitere im KI-LiveS-Projekt genutzte Datenquellen	14
4 Bewertung der Datenquellen	15
4.1 Datenquellenbezogene Test- und Evaluierungskriterien.....	15
4.2 Priorisierung der Nutzung von Datenquellen.....	18
4.2.1 Auswahl einer Methode zur Priorisierung von potenziellen Datenquellen	18
4.2.2 Anwendungsschema für die Priorisierungsmethode der Nutzwertanalyse.....	21
5 Evaluation potenzieller Datenquellen.....	23
6 Ausblick auf weitere Datenquellen für ein „Data Repository“ im Verlauf des KI-LiveS-Projekts	25
7 Möglichkeiten für ein Data Repository im KI-LiveS-Projekt.....	27
Literaturverzeichnis.....	30
Anhang A: Leitfaden Experteninterview.....	36
Anhang B: Ausgefüllte Leitfäden zur Datenerfassung im KI-LiveS-Projekt.....	39

Abkürzungs- und Akronymverzeichnis

AAP	Automatic Aural Procedure
ABC-D	Activities-Specific Balance Confidence Scale – German version auch: Activities-Specific Balance Confidence-Skala
ACI	Acoustic Complexity Index
a. d.	an der
ADI	Acoustic Diversity Index
AEI	Acoustic Evenness Index
AG	Aktiengesellschaft
AHP	Analytic Hierarchical Process, auch: Analytic Hierarchy Process
AI	Artificial Intelligence
AKT	Alters-Konzentrations-Test
AR	Acoustic Richness
Aufl.	Auflage
AVI	Audio Video Interleave
BDI-II	Beck depression inventory–II
BIO	Bioacoustic Index
BMBF	Bundesministerium für Bildung und Forschung
bzw.	beziehungsweise
ca.	circa
CBR	Case-based Reasoning
CG	Computergraphik
CERAD	Consortium to Establish a Registry for Alzheimer’s Disease
CMS	Crane Management System
cm	Zentimeter
CPSM	Communicating Business Process and Software Models
Co. KG	Compagnie Kommanditgesellschaft
CSV	Comma-separated Values
CT	Computertomographie
CVPR	Conference on Computer Vision and Pattern Recognition
DAP	Direct Aural Procedure
d. h.	das heißt
DICOM	Digital Imaging and Communications in Medicine
Dipl.	Diplom
DemTec	Dementia Toolkit for Effective Communication
DLR	Deutsches Zentrum für Luft- und Raumfahrt e. V.
DNA	desoxyribonucleic acid

DOC	Document
DOC(X)	Document XML
Doz.	Dozent
Dr.	Doktor
DSGVO	Datenschutz-Grundverordnung
EATWOS	Efficiency Analysis Technique With Output Satisficing
EDV	Elektronische Datenverarbeitung
EKG	Elektrokardiogramm
E-Mail	Electronic Mail
ES	eingebettete Systeme
et al.	et alii
EU	Europäische Union
e. V.	eingetragener Verein
IEEE	Institute of Electrical and Electronics Engineers
f.	folgende
ff.	fortfolgende
FLAC	Free Lossless Audio Codec
FLAIR	fluid attenuated inversion recovery
fMRT	funktionelle Magnetresonanztomographie
FPI-R	Freiburger Persönlichkeitsinventar
FWT	Farb-Wort-Interferenztest
FZ	Forschungszentrum
Gbyte	Gigabyte
GFOS mbH	Gesellschaft für Organisationsberatung und Softwareentwicklung mit beschränkter Haftung
gGmbH	gemeinnützige Gesellschaft mit beschränkter Haftung
GmbH	Gesellschaft mit beschränkter Haftung
H	Acoustic Entropy Index
Hf	Spectral Entropy Index
HNR	Heinz Nixdorf Recall
Hrsg.	Herausgeber
Ht	Temporal Entropy Index
HTML	Hypertext Markup Language
Hz	Hertz
ID	Identification
i. d. R.	in der Regel
IMIBE	Institut für Medizinische Informatik, Biometrie und Epidemiologie

Ing.	Ingenieur
inkl.	inklusive
insb.	insbesondere
InUPH	Institut für Urban Public Health
ISSN	International Standard Serial Number
IT.NRW	Information und Technik Nordrhein-Westfalen
jCORa	java based Case- and Ontology-based Reasoning Application
Jg.	Jahrgang
JPEG	Joint Photographic Experts Group
KG	Kommanditgesellschaft
KI	Künstliche Intelligenz
KI-LiveS	KI-Labor für verteilte und eingebettete Systeme
KOWIEN	Kooperatives Wissensmanagement in Engineering-Netzwerken
kW	Kilowatt
LEAP-Q	The Language Experience and Proficiency Questionnaire
LPS	Lipopolysaccharid
M	Amplitude Index
m	Meter
MByte	Megabyte
MGS	MehrGenerationenStudie
min	Minute
MKV	Dateiendung für Dateien im Matroska-Format
MPEG	Moving Picture Experts Group
MP3	Dateiendung für Dateien im Format MPEG-1 Audio Layer 3
MP4	Dateiendung für Dateien im Format MPEG-4, Part 14
MRT	Magnetresonanztomographie
M. Sc.	Master of Science
MS Project	Microsoft Project
MWh	Megawattstunde
NDSI	Normalized Difference Soundscape Index
No.	Number
NP	Number of Peaks
Nr.	Nummer
NWA	Nutzwertanalyse
o. Ä.	oder Ähnliches
OCRA	Operational Competitiveness Rating
OrGoLo	Organisatorische Innovationen mit Good Governance in Logistik-Netzwerken

OWA	Ordered Weighted Average
Ök.	Ökonomin
PAW	Pawaa File
PDF	Portable Document Format
PIM	Institut für Produktion und Industrielles Informationsmanagement
PNG	Portable Network Graphics
PPT(X)	PowerPoint (neues Dokument-Format)
Priv.	Privat
RAW	Format für Audio-Dateien (storing uncompressed audio in raw form)
RCA	Ruhr Campus Academy
rer. nat.	rerum naturalium
s. u.	siehe unten
S.	Seite
SALVE	Acoustic Quality and Health in Urban Environments
sec	Sekunde
SKID 1+2	Strukturiertes Klinisches Interview für DSM-IV
SNP chips	Single Nucleotide Polymorphism
SQL	Structured Query Language
Tel.	Telefon
TEU	Twenty-foot Equivalent Unit
TIFF	Tagged Image File Format
TMT	Trail Making Test
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TU DO	Technische Universität Dortmund
TuL	Transportsysteme und -logistik
txt	Text
u.	und
u. a.	unter anderem
UDE	Universität Duisburg-Essen
UKE	Universitätsklinikum Essen
URL	Uniform Resource Locator
vgl.	vergleiche
Vol.	Volume
VPN	Virtual Private Network
VS	verteilte Systeme
WAV	Waveform Audio File
WLC	Weighted Linear Combination

WST Wortschatztest
xlsx Excel Spreadsheet XML
z. B. zum Beispiel

Tabellenverzeichnis

	<u>Seite</u>
Tabelle 1: Kennzahlen des Projekts SALVE	10
Tabelle 2: Anwendungschema zur Nutzwertanalyse für Testzwecke	21
Tabelle 3: Anwendungschema zur Nutzwertanalyse für Evaluierungszwecke	22

1 Bedeutung von Datenquellen für das KI-LiveS-Projekt

Dieser Projektbericht fokussiert die Erschließung von Datenquellen, die im Rahmen des KI-LiveS-Projekts genutzt werden können, um ein Data Repository¹ aufzubauen. Dieser Analysefokus bildet aus wissenschaftlicher Perspektive den „Gegenstandsbereich“ der hier vorgelegten Untersuchungen.² Die Erschließung von Datenquellen nimmt im Rahmen des KI-LiveS-Projekts eine zentrale Stellung ein.

Das Ziel eines solchen Data Repository ist es, in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für Entwicklungsarbeiten im KI-Umfeld bereitzustellen. Diese Entwicklungsarbeiten umfassen – unter Berücksichtigung der verschiedenen Forschungsschwerpunkte³ innerhalb des Projektkonsortiums des KI-LiveS-Projekts – insbesondere die Entwicklung von datengetriebenen KI-Tools⁴.

Vor allem soll das Data Repository reale oder zumindest realitätsnahe (z. B. anonymisierte) Daten zur Verfügung stellen, die sich für Tests und Evaluierungen von KI-Tools nutzen lassen. Im Vordergrund stehen Tests und Evaluierungen derjenigen KI-Tools, die im KI-Labor des KI-LiveS-Projekts entwickelt – d. h. konzipiert und implementiert – werden, um den Transfer von Erkenntnissen und Ergebnissen⁵ der KI-Forschung in die betriebliche Praxis zu unterstützen. Darüber hinaus soll ein solches Data Repository langfristig – über das Ende der Förderung des KI-LiveS-Projekts hinaus – im Interesse der nachhaltigen Weiternutzung der Projektarbeiten dazu dienen, Daten für Tests und Evaluierungen von KI-Tools im betrieblichen Einsatz zur Verfügung zu stellen.⁶

Neben der Erschließung von Datenquellen wird in diesem Projektbericht zudem skizziert, welche Möglichkeiten bestehen, ein Data Repository aufzubauen.

-
- 1) Einen Überblick über verschiedene definitorische Ansätze von Repositorien im forschungsbezogenen Kontext bietet z. B. ASCHENBRENNER/NEUROTH (2011), S. 102 f.
 - 2) Dieser Gegenstandsbereich betrifft zunächst vor allem das Teilprojekt „KI-Brainwareentwicklung: intelligente, computergestützte Wiederverwendung von Erfahrungswissen im Projektmanagement“, das vom Institut für Produktion und Industrielles Informationsmanagement im Rahmen des Gesamtprojekts „KI-LiveS“ verantwortet wird. Neben der Fokussierung auf Datenquellen für das Teilprojekt „KI-Brainwareentwicklung“ finden sich in diesem Projektbericht auch weitere Datenquellen, die für andere Teilprojekte des KI-LiveS-Projekts von Interesse sind. Für spätere Überarbeitungen (Auflagen) des Projektberichts Nr. 2 soll der Gegenstandsbereich je nach Bedarf auf weitere Teilprojekte des KI-LiveS-Projekts erweitert werden.
 - 3) Auf die verschiedenen Forschungsschwerpunkte innerhalb des Projektkonsortiums des KI-LiveS-Projekts wird an späterer Stelle ausführlicher eingegangen.
 - 4) Unter KI-Tools werden in diesem Projektbericht primär Softwareprodukte (synonym: Software, Programme, Anwendungen, Applikationen, Systeme oder „Lösungen“) verstanden, in deren Entwicklung Erkenntnisse aus der Erforschung Künstlicher Intelligenz (KI) eingeflossen sind. Sekundär umfasst der Begriff der KI-Tools aber auch dezidierte Hardware, die auf spezielle Leistungsanforderungen von KI-Software zugeschnitten ist, sowie entsprechende Hard- und Software-Kombinationen.
 - 5) Mit Ergebnissen der KI-Forschung sind in diesem Projektbericht KI-Tools, d. h. KI-bezogene Hard- oder Software, gemeint.
 - 6) In welchem Ausmaß zu diesem nachhaltigen Weiternutzungsziel seitens des KI-LiveS-Projekts tatsächlich beigetragen werden kann, wird im vorliegenden Projektbericht insbesondere im Hinblick auf die Bereitschaft von Unternehmen, ihre test- und evaluierungsrelevanten Daten in einem öffentlich zugänglichen Data Repository bereitzustellen, kritisch erörtert werden.

Das grobe Vorgehen der Datenquellenanalyse im Hinblick auf ein Data Repository, das im Rahmen des KI-LiveS-Projekts angestrebt wird, lässt sich in acht Phasen unterteilen:

- Identifizierung von potenziellen Datenquellen im eingangs spezifizierten Gegenstandsreich, d. h. von Datenquellen zum Aufbau eines Data Repository insbesondere für Test- und Evaluierungszwecke im Hinblick auf KI-Tools,
- Vorstellung von Erhebungsmethoden für potenzielle Datenquellen einschließlich einer begründeten Methodenauswahl,
- Erhebung von potenziellen Datenquellen,
- Aufstellung eines Bewertungsframeworks für die potenziellen Datenquellen,
- Bewertung der erhobenen Datenquellen,
- Evaluation der Datenquellenerhebung,
- Ausblick auf weitere Datenquellen, die bis zum Ende des KI-LiveS-Projekts noch erschlossen werden können, sowie
- Analyse der Möglichkeiten für den Aufbau eines Data Repository im KI-LiveS-Projekt.

Diese acht Phasen werden in den nachfolgenden Kapiteln jeweils näher beschrieben.

2 Erhebung potenzieller Datenquellen

2.1 Identifizierung potenzieller Datenquellen

Als Voraussetzung für die Erhebung von Datenquellen gilt es zunächst, potenzielle Datenquellen für das intendierte Data Repository zu identifizieren. Das Identifizieren potenzieller Datenquellen orientiert sich an den verschiedenen Forschungsschwerpunkten einzelner universitärer Projektpartner innerhalb des Projektkonstortiums des KI-LiveS-Projekts.

Im KI-Labor sollen „datengetriebene“ Anwendungsbereiche von KI-Tools besonders intensiv betreut werden. Dazu gehören beispielsweise: Sensordaten aus der Instandhaltung von Containerbrücken, Videodaten aus intralogistischen Systemen, Sensordaten zum Monitoring von medizinischen Implant-Kits, Videodaten von MRT-Aufnahmen, Audiodaten aus urbanen Umgebungen sowie Handlungs-, Kunden-, Lieferanten- und Produktdaten im Projektmanagement. Als potenzielle Datenquellen bieten sich alle Datenquellen an, die Daten vorhalten, die sich vor allem für die vorgenannten Forschungsschwerpunkte nutzen lassen.

Beispielsweise umfassen solche Datenquellen für das Teilprojekt „KI-Brainwareentwicklung“, das sich auf die „intelligente“ Wiederverwendung von Erfahrungswissen im Projektmanagement konzentriert, zunächst alle denkmöglichen Quellen, in denen Erfahrungswissen (grob gesprochen: Daten)⁷ über das Management bereits durchgeführter Projekte gespeichert werden können. Dieses Erfahrungswissen soll für Tests und Evaluierungen des ontologiegestützten Case-based-Reasoning-Systems (CBR-System)⁸ jCORA⁹ genutzt werden. Dieses KI-Tool wurde am Institut für Produktion und Industrielles Produktionsmanagement der Universität Duisburg-Essen für die „intelligente“ Wiederverwendung von Erfahrungswissen im Projektmanagement entwickelt und wird im Rahmen des KI-LiveS-Projekts im Sinne des Transferziels dieses Projekts vor allem im Hinblick auf eine größere Benutzerfreundlichkeit und – hinsichtlich verschiedenartiger Projektarten sowie Branchen – breitere Anwendbarkeit verbessert.

7) In diesem Projektbericht werden die Bezeichnungen „Daten“, „Informationen“ und „Wissen“ in stark vergrößernder Weise synonym behandelt, weil es im hier betrachteten Kontext von „Datenquellen“ auf die durchaus möglichen begrifflichen Nuancierungen nicht ankommt. Beispielsweise könnte zwischen Daten als rein *syntaktisch* definierten Symbolclustern, Informationen als inhaltlich interpretierten (also „*semantisch* angereicherten“) Daten und Wissen als komplexen Informationszusammenhängen, die sich grundsätzlich zur Lösung von Problemen eignen (und somit eine *pragmatische* Anwendungsdimension erschließen), unterschieden werden. Aber eine solche Unterscheidung führt in diesem Projektbericht nicht weiter, weil zunächst von gespeicherten *Daten* in z. B. Datenbanken („Datenquellen“) auszugehen ist, deren inhaltliche Interpretation und deren Anwendung zur Problemlösung noch nicht feststeht, solange diese Daten in einer Datenquelle vorgehalten werden. Erst die später mögliche Interpretation dieser Daten beispielsweise als Bestandteile einer Fallbeschreibung, eines Fallresultats oder einer Fallbewertung, wie sie im Rahmen des Case-based Reasonings (CBR) üblich ist, verleiht ihnen den Charakter als *Informationen*. Des Weiteren kann der komplexe Zusammenhang vielfältiger Informationen, die über einen alten Fall in der Triple-Struktur von Fallbeschreibung, Fallresultat und Fallbewertung vorliegt, den Charakter von (Erfahrungs-)Wissen erlangen, das zur Lösung neuer Probleme (im Case-based Reasoning synonym: neuer Fälle) angewendet wird. Den Daten, die in einer Datenquelle gespeichert sind (syntaktische Dimension), kann aber nicht „angesehen“ werden, ob sie später inhaltlich interpretiert werden (semantische Dimension) oder sogar zur Lösung eines Problems angewendet werden (pragmatische Dimension). Wegen dieser *Unbestimmtheit* des später möglichen Umgangs mit Daten wird in diesem Projektbericht vergrößernd von der voranstehend skizzierten Möglichkeit, zwischen Daten-, Informations- und Wissenbegriff zu differenzieren, der Einfachheit halber abstrahiert. Diese terminologische Vereinfachung erlaubt es, im selben Kontext eines KI-Tools beispielsweise sowohl von „Datenquellen“ als auch von „Erfahrungswissen“ zu sprechen, ohne aufwändig zwischen Daten und Wissen unterscheiden zu müssen.

8) CBR-Systeme werden in diesem Projektbericht als ein Spezialfall von KI-Tools behandelt und daher auch als KI-Tools angesprochen.

9) Vgl. BERGENRODT/KOWALSKI/ZELEWSKI (2015), insbesondere S. 512 ff.; BERGENRODT/KOWALSKI (2015), insbesondere S. 34 ff.

Erfahrungswissen sollte im Rahmen von Aufgaben des Projektmanagements die folgende dreiteilige Struktur aufweisen:¹⁰

- Projektbeschreibung im Sinne der Beschreibung eines Projektauftrags (z. B. einer „Ausschreibung“),
- Projektergebnis im Sinne der Ergebnisse, die im Verlauf der Projektdurchführung erarbeitet wurden, sowie
- Projektbewertung in der Gestalt von Urteilen über betriebswirtschaftlich besonders relevante Aspekte von Projektbeschreibung und Projektergebnis, vor allem im Hinblick auf „kritische Erfolgsfaktoren“, die wesentlich zum Gelingen oder Mislingen des Projekts beigetragen haben.

In der Praxis des alltäglichen Projektmanagements werden verschiedene „generische“ Formen zur Dokumentation von projektbezogenem Erfahrungswissen eingesetzt. Zu den gängigsten Dokumentationsformen von Erfahrungswissen gehören Lessons Learned¹¹, Debriefings¹², Reports¹³ und Post Project Reviews¹⁴. Darüber hinaus wird projektbezogenes Erfahrungswissen auch in spezieller Projektmanagement-Software gespeichert.¹⁵ Zu den gängigen Dateiformaten zur Speicherung von projektbezogenem Erfahrungswissen dienen DOC(X)-, PPT(X)- und PDF-Dateien.¹⁶ Vereinzelt werden auch Excel- oder HTML-Dateien verwendet.

Projektbezogenes Erfahrungswissen kann zudem in den Köpfen von Mitarbeitern „eingeschlossen“ sein. Dies betrifft im Besonderen das sogenannte implizite Wissen¹⁷. Implizites Wissen im Kontext des Projektmanagements umfasst z. B. nicht dokumentierte Handlungs- und Lösungsstrategien, individuelle Erfahrungen sowie kreative Lösungsansätze der involvierten Mitarbeiter. Oftmals sind sich Mitarbeiter über das Ausmaß ihres Wissensschatzes jedoch nicht vollumfänglich bewusst, sodass für ein Unternehmen der Umfang und der Inhalt des impliziten Wissens seiner Mitarbeiter zumindest partiell unbekannt bleibt.

10) Die nachfolgende Strukturierung in Projektbeschreibung, Projektergebnis (oftmals synonym: Projektlösung) und Projektbewertung folgt der Strukturierung konventioneller CBR-Systeme, die oftmals zwischen Fallbeschreibung, Fallresultat und Fallbewertung unterscheidet; vgl. zur konventionellen Konstitution von CBR-Systemen BEIERLE/KERN-ISBERNER (2019), S. 173 f.

11) Vgl. z. B. JUGDEV (2012), S. 13 ff.; FÜRSTENAU/LANGFERMANN/BORN et al. (2005), S. 1028.

12) Vgl. z. B. SCHINDLER/EPPLER (2003), S. 220 ff.

13) Vgl. z. B. FREITAG (2011), S. 116.

14) Vgl. z. B. BUSBY (1999), S. 23 ff.; DANIEL/KUNZ/WAHNIG (2009), S. 183 ff.

15) In der betrieblichen Praxis existiert eine Vielzahl von Projektmanagement-Softwares, in denen Erfahrungswissen gespeichert werden kann. Gängige, in der betrieblichen Praxis verwendete Projektmanagement-Softwares sind z. B. MS Project, Jira und Windchill.

16) Vgl. zur Verwendung der genannten Dateiformate vor allem die empirischen Erhebungen, über die in SCHAGEN/ZELEWSKI/HEEB (2020), S. 35 f., beruhend auf einer Befragung von Projektmanagern, berichtet wird. Die Quellenangabe gilt auch für den nachfolgenden Satz.

17) Die Unterscheidung zwischen explizitem und implizitem Erfahrungswissen lehnt sich an die Unterscheidung zwischen explizitem und implizitem Wissen an, die auf die Arbeit von POLANYI zurückgeht (vgl. POLANYI (2009), S. 4 ff.) und in der einschlägigen Fachliteratur zum Wissensmanagement intensiv rezipiert wird. Zwar lässt sich über die oftmals vorgenommene „strikte“ Unterscheidung zwischen explizitem und implizitem Wissen streiten. Dies liegt einerseits an Möglichkeiten zur Explizierung impliziten Wissens, die beispielsweise durch die sogenannte „Wissensspirale“ angedeutet werden; vgl. dazu NONAKA/TAKEUCHI (2012), S. 72 ff. Andererseits lässt sich ebenso an die Nutzung von Inferenzregeln denken, die es in formalsprachlichen Argumentationskontexten gestatten, implizites Wissen, das in der Formulierung einer Theorie oder eines Modells (als „Miniatur-Theorie“) verborgen ist, zu explizieren. Jedoch werden Vorbehalte der zuvor skizzierten Art nicht weiter berücksichtigt, um die Anschlussfähigkeit der Ausführungen an die einschlägige Fachliteratur zu wahren.

2.2 Vorstellung der Erhebungsmethode

Das Erheben potenzieller Datenquellen im Rahmen des KI-LiveS-Projekts erfolgt anhand eines Leitfadens. Dieser Leitfaden zielt darauf ab, die Bereitschaft zur Bereitstellung von Daten als Datenquellen für ein Data Repository im Rahmen des KI-LiveS-Projekts zu erkunden. Der Leitfaden ist als ausfüllbares PDF-Formular konstruiert. Dies erlaubt eine rasche Verbreitung des Leitfadens seitens der Praxis- und Universitätspartner des KI-LiveS-Projekts via E-Mail und auch eine schnelle computergestützte Verarbeitung der ausgefüllten sowie via E-Mail zurückgesandten PDF-Formulare. Deshalb wird die „elektronische“ Handhabung des Leitfadens als PDF-Formular beispielsweise gegenüber dem Ausdruck und postalischen Zusenden des Leitfadens vorgezogen.

Der Leitfaden wurde zwar unter der Federführung des Instituts für Produktion und Industrielles Informationsmanagement der Universität Duisburg-Essen erstellt, beschränkt sich jedoch nicht nur auf Datenquellen für das Teilprojekt „KI-Brainwareentwicklung“. Stattdessen dient der Leitfaden zur umfassenden Erhebung aller Datenquellen, die für das Gesamtprojekt „KI-LiveS“ von Bedeutung sein könnten.

Zu Beginn des Leitfadens wird der Untersuchungsgegenstand erläutert.¹⁸ Insbesondere wird auf die Bedeutung eines Data Repository hingewiesen. Zugleich wird Verständnis für die Sensibilität unternehmensinterner Daten geäußert. Die nachfolgenden Fragen lassen sich grob in sechs Kategorien unterteilen, die im Verlauf dieses Kapitels näher erläutert werden:

- Bereitschaft zur Datenbereitstellung,
- allgemeine Angaben zu den Daten,
- technische Daten zu den Angaben,
- Strukturierung der Daten,
- Quantität der Daten sowie
- Qualität der Daten.

Zunächst erfolgt die Frage, ob grundsätzlich die Bereitschaft besteht, Datenquellen zugänglich zu machen, um die dort vorgehaltenen Daten für die Nutzung durch Dritte zur Verfügung zu stellen. Diese Frage kann entweder bejaht oder verneint werden. Wird diese Frage verneint, endet das Ausfüllen des Leitfadens an dieser Stelle. Wird diese Frage bejaht, erfolgt das Erfragen allgemeiner Angaben zu den Daten. Diese Angaben umfassen das Datum der Erfassung und die Herkunft der Daten sowie die Möglichkeit, mit Schlüsselwörtern die vorgehaltenen Daten zu charakterisieren. Bezüglich der Herkunft der Daten kann zwischen den Items „öffentliche Daten“, „Unternehmensdaten“, „private Daten“ und „Sonstiges“ gewählt werden.

Anschließend werden technische Angaben zu den Daten abgefragt. Zunächst wird um eine Einordnung der Daten in Kategorien gebeten. Als Kategorien sind „Sensordaten“, „Tabellendaten“, „Textdaten“ und „Sonstiges“ vorgegeben. Sensordaten werden unterteilt in die Items „Audiiodaten“, „Bilddaten“ und „Videodaten“. Textdaten werden unterteilt in die Items „reiner Text“, „Text mit eingebetteten Grafiken“, „Text mit eingebetteten Videodaten“ und „Text mit eingebetteten Audiiodaten“. Nach dieser Einordnung der Daten wird erfragt, ob die Daten in einem standardisierten Format vorliegen. Dabei wird zwischen den Kategorien „Bilddaten“, „Videodaten“, „Audiiodaten“, „Tabellendaten“, „Textdaten“ und „weiteren Datenformaten“ unterschieden. Im Einzelnen kann zunächst zwischen folgenden weit verbreiteten Datenformaten ausgewählt werden:

18) Der vollständige Leitfaden kann dem Anhang A dieses Projektberichts entnommen werden. Die nachfolgende Skizzierung des Leitfadens erfasst die abgefragten Items nicht vollständig, sondern bietet nur einen exemplarischen Einblick in die Leitfadengestaltung. Für eine vollständige Darstellung der Items wird auf den bereits angesprochenen Anhang A verwiesen.

- für Bilddaten zwischen den Datenformaten PNG, JPEG, TIFF, PAW und DICOM,
- für Videodaten zwischen den Datenformaten AVI, MPEG/MP4 und MKV,
- für Audiodaten zwischen den Datenformaten WAV, MP3, OGG, FLAC und RAW,
- für Tabellendaten zwischen den Datenformaten Excel (XLS oder XLSX) und CSV,
- für Textdaten zwischen den Datenformaten DOC, DOCX, PDF und TXT.

Bei weiteren, seltener angewandten Datenformaten ist das Format in einem Freifeld selbst anzugeben.

Die Erhebung der Strukturierung der Daten erfolgt anhand dreier Fragen. Zunächst erfolgt die Frage, ob die Daten strukturiert sind. Dies kann entweder bejaht oder verneint werden. Zudem besteht die Möglichkeit, keine Angabe zu machen. Anschließend wird erfragt, ob die Daten für überwachte Lernverfahren geeignet sind. Dies kann wiederum bejaht oder verneint werden. Abermals besteht die Möglichkeit, auf eine Angabe zu verzichten. Die dritte Frage behandelt, ob die Problemstellung klar formuliert ist. Es kann wiederum zwischen den drei zuvor erwähnten Antwortoptionen gewählt werden.

Hinsichtlich der Quantität der Daten wird zunächst erfragt, wie groß der durchschnittliche Umfang der Dateien ist. Es erfolgt eine Unterteilung in „Bilddaten“, „Videodaten“, „Audiodaten“, „Tabellendaten“, „Textdaten“ und „Sonstiges“. Die Angabe soll anhand einer groben Schätzung in den Einheiten MByte oder GByte erfolgen. Ebenso wird erfragt, wie viele Dateien in dem jeweiligen Unternehmen für die Unterstützung der Geschäftsprozesse pro Jahr – grob geschätzt – verwendet werden. Betrachtet werden wieder die gleichen Kategorien wie bei der ersten Frage zur Datenquantität. Zur Auswahl stehen die Antwortoptionen (Items) „weniger als 100“, „zwischen 100 und weniger als 1.000“, „zwischen 1.000 und weniger als 10.000“ sowie „10.000 und mehr“.

Hinsichtlich der Qualität der Daten wird zunächst erfragt, wie hoch der Anteil annotierter Daten ist. Es erfolgt eine prozentuale Abfrage anhand der Items „0 %“, „mehr als 0 % und höchstens 25 %“, „mehr als 25 % und höchstens 50 %“, „mehr als 50 % und höchstens 75 %“ sowie „mehr als 75 %“. Zudem kann die Antwortoption „keine Aussage“ gewählt werden. Anschließend wird die Homogenität der Daten erfragt. Es soll eine Einschätzung anhand der Items „gering“, „mittel“ und „hoch“ getroffen werden. Wiederum besteht die Möglichkeit, keine Aussage zu tätigen. Als dritte Frage im Bereich der Datenqualität wird erfragt, ob die Daten bereits in irgendeiner Form verarbeitet worden sind. Betrachtet werden Sensor-, Tabellen- und Textdaten und es wird erfragt, ob die Daten entweder direkt von einem Gerät (Sensordaten), aus Tabellendateien (Tabellendaten) oder aus Textdateien (Textdaten) stammen oder ob eine Nachbearbeitung der Daten erfolgt ist. Wenn letzteres der Fall ist, wird zudem erfragt, in welcher Weise diese Nachbearbeitung der Daten durchgeführt wurde. Als letzte Frage im Bereich der Datenqualität und gleichzeitig als Abschluss des Leitfadens wird erfragt, in welche Kategorien die intendierten Anwendungen der vorgehaltenen Daten fallen. Vorgegebene Kategorien sind die Klassifikation im Sinne der Einteilung von (objektbezogenen) Daten in eine oder mehrere vorgegebene Klassen, die Segmentierung von Datengesamtheiten im Sinne der Einteilung in zusammenhängende Bereiche, die Detektion/Lokalisation im Sinne der Bestimmung relevanter Positionen und die Wiederverwendung von projektbezogenem Erfahrungswissen. Zudem wird in der Kategorie „Sonstiges“ die Möglichkeit geboten, andere Kategorien zu benennen. Mehrfachnennungen sind möglich.

2.3 Identifizierung von Stakeholdern für die Erhebung von Datenquellen

Wie eingangs begründet, fokussiert sich die Erhebung von Datenquellen im Rahmen des KI-LiveS-Projekts auf die Stakeholder, die den Gegenstandsbereich der Projektarbeiten maßgeblich prägen. Als Stakeholder werden nachfolgend vor allem¹⁹ jene Akteure angesehen, die als Kooperationspartner der betrieblichen Praxis für das Teilprojekt „KI-Brainwareentwicklung“ gewonnen werden konnten und nachfolgend als „Praxispartner“ bezeichnet werden. Hinzu kommen Akteure, die auf universitärer Seite am KI-LiveS-Projekt beteiligt sind²⁰ und nachfolgend als „Universitätspartner“ bezeichnet werden.

Als Praxispartner konnten für das Teilprojekt „KI-Brainwareentwicklung“ gewonnen werden:

- Accenture Technology Solutions GmbH (Dependance Essen, Zeche Zollverein), Ansprechpartner derzeit: Frau Wenholz, Frau Hoppe und Herr Tippelt;
- adesso AG, Ansprechpartner derzeit: Frau Dr. Carell und Herr Professor Dr. Angele;
- Atos Information Technology GmbH, Ansprechpartner derzeit: Frau Muhmann und Herr Sethupathy;
- E.Y. (Ernst & Young) Deutschland, Ansprechpartner derzeit: Frau Hinsen und Herr Dr. Alan;
- GFOS mbH, Ansprechpartner derzeit: Herr Röhrig und Frau Röhrig;
- Grunenberg & Comp. GmbH, Ansprechpartner: Herr Dr. Grunenberg, Herr Steinkötter und Herr Hülsbusch;
- OFIGO GmbH & Co. KG, Ansprechpartner derzeit: Herr Rittel und Herr van Acken;
- Ruhr Campus Academy (RCA) gGmbH, Ansprechpartner derzeit: Herr Scholz;
- thyssenkrupp Industrial Solutions AG, Ansprechpartner derzeit: Herr Dr. Ritz und Herr Göksu.

Praxispartner im Kontext weiterer Teilprojekte des Verbundprojekts KI-LiveS (verantwortlich: Fachgebiet Transportsysteme und -logistik, Universität Duisburg-Essen) sind:

- Arvato;
- Clatronic International GmbH;
- Duisburger Hafen AG (duisport);
- SDZ GmbH;
- Zolitron Technology GmbH.

19) Vgl. dazu die Öffnungsklausel in der Fußnote 2) dieses Projektberichts. Dieser Öffnungsklausel zufolge erstreckt sich das Identifizieren von potenziellen Datenquellen im Zuge dieses Projektbereichs zunächst vor allem auf das Teilprojekt „KI-Brainwareentwicklung: intelligente, computergestützte Wiederverwendung von Erfahrungswissen im Projektmanagement“, das vom Institut für Produktion und Industrielles Informationsmanagement im Rahmen des Gesamtprojekts „KI-LiveS“ verantwortet wird. Es ist beabsichtigt, dass spätere Überarbeitungen (Auflagen) des Projektberichts weitere Teilprojekte des KI-LiveS-Projekts mit weiteren Stakeholdern und ihren Datenquellen in stärkerem Maße einbeziehen.

20) Die ausdrückliche Berücksichtigung von Universitätspartnern an dieser Stelle beruht darauf, dass potenzielle Datenquellen für ein Data Repository auch aus Forschungs- oder Praxisprojekten von einzelnen Universitätspartnern hervorgehen können.

Nachfolgende Universitätspartner sind zudem am KI-LiveS-Projekt beteiligt:

- Herr Univ.-Prof. Dr.-Ing. Bernd Noche (Gesamtprojektleiter), Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften, Fachgebiet Transportsysteme und -logistik;
- Frau Univ.-Prof. Dr. rer. nat. Susanne Moebus, Universitätsklinikum Essen, Institut für Urban Public Health;
- Herr Univ.-Prof. Dr. rer. nat. Gregor Schiele, Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften, Fachgebiet Eingebettete Systeme der Informatik;
- Herr Priv.-Doz. Dr. habil. Frank Weichert
Technische Universität Dortmund, Informatik VII, Lehrstuhl für Computergraphik;
- Herr Univ.-Prof. Dr.-Ing. Torben Weis, Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften, Fachgebiet Verteilte Systeme;
- Herr Univ.-Prof. Dr. Stephan Zelewski, Universität Duisburg-Essen, Fakultät für Wirtschaftswissenschaften, Institut für Produktion und Industrielles Informationsmanagement.

2.4 Durchführung der Datenquellenerhebung

Zur Erhebung der Datenquellen wurde der in Kapitel 2.2 vorgestellte Leitfaden zur Datenerhebung als PDF-Formular an ausgewählte²¹ Praxis- und Universitätspartner versandt. Das Versenden des Leitfadens zur Datenerhebung erfolgte per E-Mail. Das PDF-Formular wurde von den angesprochenen Praxis- und Universitätspartnern ausgefüllt und an das für die Organisation der Erhebung verantwortliche Institut für Produktion und Industrielles Informationsmanagement zurückgesandt.

Die PDF-Formulare zur Datenerhebung wurden ab April 2020 versandt. Insgesamt wurden zehn PDF-Formulare an acht Praxispartner und zwei Universitätspartner versandt. Die Rücklaufquote bei den Praxispartnern betrug 25 %²², bei den Universitätspartnern 100 %.

21) Zu den ausgewählten Praxispartnern gehörten zunächst – bis auf eine Ausnahme – alle Praxispartner des Teilprojekts „KI-Brainwareentwicklung“, die in Kapitel 2.3 genannt wurden. Ausgenommen war die Ruhr Campus gGmbH, die als Dienstleistungsunternehmen im Bereich der akademischen Weiterbildung für die Erhebung potenzieller Datenquellen nicht in Betracht kam. Zudem wurde der Leitfaden an zwei Universitätspartner, das Fachgebiet Transportsysteme und -logistik (Universität Duisburg-Essen, Fakultät für Ingenieurwissenschaften) sowie das Institut für Urban Public Health (Universitätsklinikum Essen), versandt. Die Versendung an das Fachgebiet Transportsysteme und -logistik erfolgte, um das Ausfüllen der PDF-Formulare seitens derjenigen Praxispartner zu koordinieren, die vom Fachgebiet Transportsysteme und -logistik betreut werden.

22) Hinsichtlich dieser Rücklaufquote ist zu berücksichtigen, dass die meisten Praxispartner auf das Ausfüllen der PDF-Formulare verzichtet haben, da in bilateralen Gesprächen zwischen den betroffenen Praxispartnern und dem Institut für Produktion und Industrielles Informationsmanagement bereits frühzeitig signalisiert wurde, dass seitens der Praxispartner keine Bereitschaft besteht, Datenquellen mit projektspezifischem Erfahrungswissen für ein Data Repository zur Verfügung zu stellen.

3 Auswertung der Datenquellenerhebung

3.1 Übersicht über potenzielle Datenquellen

Hinsichtlich der Erhebung *potenzieller* Datenquellen für *projektbezogenes Erfahrungswissen* ergibt sich aus der Auswertung der ausgefüllten und zurückgesandten PDF-Formulare²³ der Praxispartner des Teilprojekts „KI-Brainwareentwicklung“, dass keine Datenquellen mit projektbezogenem Erfahrungswissen jeglicher Art dem Projektkonsortium des KI-LiveS-Projekts für ein Data Repository zur Verfügung gestellt werden.

Aus vertiefenden Gesprächen zwischen Mitgliedern des Instituts für Produktion und Industrielles Informationsmanagement einerseits und den befragten Praxispartnern andererseits ging hervor, dass die strikte Zurückhaltung des Praxispartner hinsichtlich einer Zurverfügungstellung ihrer Daten zu projektbezogenem Erfahrungswissen vor allem auf zwei Gründen beruht. Erstens bestehen grundsätzliche Bedenken gegenüber der „Preisgabe“ von projektbezogenem Erfahrungswissen, das in einem semantischen Umfeld von Begriffen wie „Betriebsgeheimnis“, „strategisches Asset“ und „immaterieller Vermögenswert“ verortet wird. Aus diesem Grund wird die Offenlegung dieses Wissens in einem Data Repository, auf das Dritte – sei es nur im Rahmen des KI-LiveS-Projekts oder auch als öffentlich zugängliche Datenquelle – zugreifen können, kategorisch ausgeschlossen. Zweitens wird von einigen der befragten Praxispartner des Teilprojekts „KI-Brainwareentwicklung“ in unterschiedlicher Deutlichkeit darauf hingewiesen, dass das projektbezogene Erfahrungswissen im Unternehmen „diffus verteilt“ sei bis hin zur Eingeschlossenheit „in den Köpfen von Projektmitarbeitern“ als implizites Wissen. Deshalb sehen sich die Unternehmen außerstande, die erheblichen Ressourcen zu investieren, die für eine Erschließung des projektbezogenen Erfahrungswissens in einem Data Repository erforderlich wären.

Hinsichtlich *weiterer potenzieller Datenquellen* für ein Data Repository im Rahmen des KI-LiveS-Projekts ergibt sich ein positiveres Bild.

Einerseits konnten vom Institut für Urban Public Health, Universitätsklinikum Essen, nachfolgende Datenquellen²⁴ erschlossen werden, die sich für ein Data Repository verwenden lassen:

- ca. 10.000 GByte Sounddaten sowie
- ca. 10 GByte Tabellendaten.

Darüber hinaus konnten weitere Datenquellen²⁵ des Instituts für Urban Public Health, Universitätsklinikum Essen, identifiziert werden, die sich zwar für Forschungszwecke grundsätzlich – also vorbehaltlich weiterer Detailvereinbarungen im Einzelfall zu Forschungszwecken, Einhaltung von Datenschutzanforderungen usw. – nutzen lassen, jedoch für ein öffentlich zugängliches Data Repository nicht zur Verfügung stehen.

Andererseits konnten vom Praxispartner Duisburger Hafen AG (duisport) umfangreiche Datenquellen²⁶ erschlossen werden. Für diese Datenquellen gilt ebenso, dass sie zwar für Forschungszwecke genutzt werden können, jedoch für ein öffentlich zugängliches Data Repository nicht zur Verfügung stehen.

23) Ein Überblick über die ausgefüllten und zurückgesandten PDF-Formulare kann dem Anhang B dieses Projektberichts entnommen werden.

24) Eine konkretisierende Beschreibung der Herkunft und Ausgestaltung dieser Datenquellen erfolgt in Kapitel 3.2.

25) Eine konkretisierende Beschreibung der Herkunft und Ausgestaltung dieser Datenquellen erfolgt in Kapitel 3.2.

26) Eine konkretisierende Beschreibung der Herkunft und Ausgestaltung dieser Datenquellen erfolgt in Kapitel 3.2.

3.2 Beschreibung der potenziellen Datenquellen

Nachfolgend werden die potenziellen Datenquellen beschrieben, die im KI-LiveS-Projekt identifiziert wurden. Sie umfassen sowohl diejenigen Datenquellen, die für ein Data Repository tatsächlich zur Verfügung stehen (erschlossene Datenquellen), als auch diejenigen Datenquellen, für die zurzeit keine Freigabe für eine Integration in ein Data Repository vorliegt (identifizierte potenzielle, aber derzeit nicht verfügbare Datenquellen).

(1) Beschreibung der Datenquellen des Instituts für Urban Public Health (InUPH)²⁷

acoustic quality and health in urban environments (SALVE) Project

Die Ende 2018 gestartete Pilotstudie SALVE ist eine interdisziplinäre Studie, deren Hauptziel die Beschreibung der urbanen akustischen Umgebung der Stadt Bochum ist. SALVE zielt auf die Erstellung eines qualitätsgesicherten, räumlich und zeitlich hochaufgelösten Datensatzes als Voraussetzung für die Untersuchung des Zusammenhangs zwischen der urbanen akustischen Umwelt, der bebauten Umwelt und der menschlichen Gesundheit ab.

Daten:

Vom 18.03.2019 bis zum 18.05.2020 wurde die akustische Umgebung an 758 Orten in Bochum aufgezeichnet. Die Messorte wurden mittels einer disproportional geschichteten Zufallsstichprobe aus den Flächennutzungstypen der Stadt bestimmt. Die Messungen erfolgten zum einen manuell (Direct Aural Procedure (DAP)) und zum anderen mittels automatischer Aufnahmegeräte (Automatic Aural Procedure (AAP)), die an ausgesuchten Bäumen angebracht wurden. Mittels DAP waren 2.920 fünfminütige Aufnahmen geplant (an 730 Orten, viermal über das Jahr verteilt). Tatsächlich konnten davon 2.433 Aufnahmen realisiert werden (24 bit depth, 48.000 Hz). Mittels AAP waren 438.000 dreiminütige Aufnahmen geplant (an 24 Orten, eine Aufnahme alle 26 Minuten). Tatsächlich konnten davon 416.797 Aufnahmen realisiert werden (16 bit depth, 44.100 Hz):

Methode	Anzahl Aufnahmen (n)	Aufnahmelänge/Einzelaufnahme	Total
DAP	2.433	5 min	12.165 min
AAP	416.797	3 min	1.250.391 min

Tabelle 1: Kennzahlen des Projekts SALVE

Alle Aufnahmen liegen im WAVE-Format vor. Zusätzlich wurden für alle Aufnahmen Lautstärke (dBa_mean, -max & -min) sowie weitere Indizes aus der Soundscape Ecology berechnet, die verschiedene Eigenschaften des Frequenzspektrums parametrisieren (ACI, ADI, AEI, AR, BIO, H, Hf, Ht, M, NDSI, NP).²⁸ Zu jeder Aufnahme existieren zudem Variablen zur Bestimmung von Zeit und Ort der Aufnahme, zu Eigenschaften der bebauten Umgebung, z. B. in Form von Landnutzungstypen sowie Wetterdaten, wenn auch nicht punktgenau. Die Daten können in verschiedenen Tabellenformaten ausgegeben werden (XLSX, CSV o. Ä.). Diese Tabellen haben insgesamt eine Größe von ca. 1 GByte. Alle Daten wurden bezüglich ihrer Plausibilität geprüft.

27) Die Beschreibung der Datenquellen des Instituts für Urban Public Health (InUPH) wurde von Herrn TIMO HASELHOFF, wissenschaftlichem Mitarbeiter am InUPH, vorgenommen.

28) Vgl. SUEUR (2018), S. 483.

Verfügbarkeit:

Alle Indizes könnten als Open Data zur Verfügung gestellt werden. Die Nutzung der Daten bedarf dann nur der Angabe der Quelle (SALVE-Projekt). Aus Gründen des Datenschutzes können die Daten allerdings nicht im WAVE-Format öffentlich zur Verfügung gestellt werden, da Stimmen möglicherweise identifiziert werden können. Diese Daten müssten mit Hilfe eines Datentransferantrags beim Institut für Urban Public Health angefragt werden.

Studie Heinz Nixdorf Recall (HNR), nachgeschaltete MehrGenerationenStudie (MGS) und 1.000Gehirne-Studie***HNR (<https://www.uni-due.de/recall-studie/>)***

Am Universitätsklinikum Essen wird seit dem Jahr 2000 die populationsbasierte Heinz-Nixdorf-Recall-Langzeitstudie mit 4.814 zufällig ausgewählten Bürgerinnen und Bürgern der Städte Bochum, Essen und Mülheim a. d. Ruhr im Alter von 45 bis 75 Jahren durchgeführt.²⁹ Die Studie umfasst im Abstand von jeweils fünf Jahren drei umfangreiche medizinisch-klinische Untersuchungen, Befragungen sowie die Erfassung umweltbezogener Expositionen (Lärm, Feinstaub) (s. u. „Daten“). Darüber hinaus werden jährliche Befragungen zur Erfassung des Gesundheitsstatus und insbesondere gesundheitlicher Outcomes (Herzinfarkt, Schlaganfall, Todesfälle usw.) durchgeführt. Das fragebogengestützte Follow-up der HNR-Studie läuft seit 20 Jahren.

Daten:

Herz-CT, Ultraschalluntersuchungen Herz und Halsschlagader; Blutdruck- und Pulsmessung; Anthropometrie: Größe, Gewicht, Taillen- und Hüftumfang; Ruhe-EKG, Blut- und Urintests incl. DNA- und SNP-Chips für genomweite Studien, Fahrradergometrie, Handkraftmessung, Riechtest, Kognitionstests, Messung schlafbezogener Atmungsstörungen, Fragebögen zu gesundheitlichen Dimensionen (Depressionen, Lebensqualität, körperliche Aktivität, Rauchen, Ernährung), Demografie, Ausbildung, Berufsbiographie, Wohnbiographie, Lebensqualität, Medikamenteneinnahme; umweltbezogene individuelle Expositionen (Grün, Lärm, Feinstaub).

Verfügbarkeit:

Die Daten der HNR-Studie werden vom Institut für Medizinische Informatik, Biometrie und Epidemiologie (IMIBE) am Universitätsklinikum Essen verwaltet. Einzelne Variablen oder Variablensätze werden nach einem vorgegebenen Prozedere für wissenschaftliche Fragestellung zur Verfügung gestellt. Daher stehen sie (bislang) nicht für ein Data Repository anderer Projekte zur Verfügung.

MGS (<https://www.uni-due.de/recall-studie/die-studien/mgs/>)

Seit dem Jahr 2013 werden auch die Partner(innen) sowie die Kinder der ersten Generation der HNR-Studienteilnehmenden nach dem Studienprotokoll der HNR-Studie in der MehrGenerationenStudie untersucht. Die Baseline-Erhebung mit umfangreichen Untersuchungen im Studienzentrum erfolgte zwischen den Jahren 2013 und 2016 mit dem Einschluss von 2.898 Personen im Alter von 18 bis 90 Jahren. Seitdem wurden eine weitere Untersuchung sowie jährliche fragebogengestützte Follow-ups durchgeführt.

Daten:

Wie HNR.

Verfügbarkeit:

Wie HNR.

29) Vgl. ERBEL/EISELE/MOEBUS et al. (2012), S. 809 ff. Das Zitat ist bis zum Absatzende gültig.

1000Gehirne-Studie

Zwischen den Jahren 2011 und 2018 wurde in Kooperation mit dem Forschungszentrum Jülich die Längsschnitt-Studie „1000Gehirne“ durchgeführt.³⁰ Die variable Alterung des Gehirns wird hier erstmals mit Langzeitdaten zu Lebensumständen, sozialen sowie genetischen Faktoren und dem Gesundheitszustand der Probanden untersucht. Das Gehirn wurde zweimal mit Abstand von 2,5 Jahren mit Hilfe struktureller und funktioneller MRT und weiterer Tests untersucht, in denen neben kognitiven Funktionen wie Gedächtnis und Aufmerksamkeit vor allem ein Schwerpunkt auf die Untersuchung der Motorik der Probanden gelegt wurde. Die beiden Rekrutierungsquellen sind die 10-Jahres-Follow-up-Kohorte der HNR sowie die Baseline-Kohorte der MGS. Von Mitte 2011 bis Mitte 2016 wurden Daten von 1.316 HNR- und MGS-Teilnehmenden in der Baseline-Untersuchung und von Mitte 2015 bis Ende 2018 Daten zu 468 Follow-up-Untersuchungen erhoben.

Daten:

Übersicht zu 3-Teslar-MRT: 3D-T1- und 3D-T2-gewichtete Scans für Strukturanalysen und Myelin-Mapping, drei Diffusions-Bildgebungssequenzen, optimiert für Diffusions-Tensor-Bildgebung, Diffusions-Bildgebung mit hoher Winkelauflösung für detaillierte Faserverfolgung und für Diffusionswölbungs-Bildgebung, Ruhezustands- und aufgabenbasierte funktionelle MRT und MR-Angiographie für die Erkennung von vaskulären Läsionen und die Kartierung von Läsionen der weißen Substanz (Protokolle: T1-MPRAGE, Diffusion 30, 60, 120 Richtungen, Resting state, T2, fMRT, FLAIR, Angio).

Übersicht zu Neuropsychologie, Motorik, Behavioural/Persönlichkeit: ABC-D, Wortschatztest (WST), Händigkeit Oldfield, Nürnberger-Alters-Alltagsaktivitäten-Skala, Lebensqualität, Frontal Behavioural Inventory-S, -A, BDI-II, FPI-R, LEAP-Q, Teil 1+2, SKID 1+2, Beanspruchung, Zufriedenheit. TMT, AKT, BKW, Benton Test, Corsi Block Tapping, Visual pattern Test, Fünf Punkte Test, LPS 50+, Abfrage BKW, Zahlennachsprechen (vorwärts), Regensburger Wortflüssigkeitstest (semantisch), Regensburger Wortflüssigkeitstest (phonematisch), CERAD, FWT (Stroop), DemTec, Wackelbrett.

Verfügbarkeit:

Die Daten werden zentral im FZ Jülich und IMIBE gespeichert. Die rohen MRT-Daten werden demnächst „open access“ zur Verfügung stehen. Darüber hinaus ist ein ähnliches Prozedere wie bei der HNR-Studie (Datennutzungsanträge) zur Datenherausgabe für wissenschaftliche Zwecke vorgesehen.

(2) Beschreibung der Datenquellen der Duisburger Hafen AG (duisport)³¹

Die Duisburger Hafen AG ist Eigentümerin und Betreiberin des Duisburger Hafens, dem größten Hinterland-Hub Europas. Mit ihren Tochtergesellschaften agiert sie weltweit als „duisport-Gruppe“ und betreibt in Duisburg und Umgebung mehrere bi- und trimodale Containerterminals, darunter die drei Terminals logport I, II und III.

Um die Betriebsdaten der Terminalkräne zu erfassen, nutzt duisport ein selbst entwickeltes „Crane Management System“ (CMS). Im CMS werden Informationen über jeden einzelnen Steuerbefehl und jeden einzelnen Betriebszustand der auf den Terminals logport I bis III betriebenen Krananlagen gespeichert. Im Rahmen des Projekts KI-LiveS wurden den Projektpartnern Leseberechtigungen für die Datensätze des CMS eingerichtet (VPN-Zugriff auf den SQL-Server). Dadurch liegen Datensätze für einen Zeitraum von November 2017 bis zum heutigen Zeitpunkt vor.

30) Vgl. CASPERS/MOEBUS/LUX et al. (2014). Das Zitat ist bis zum Absatzende gültig.

31) Die Beschreibung der Datenquellen von duisport wurde von Frau SANDRA SCHMITZ, wissenschaftliche Mitarbeiterin am Lehrstuhl für Transportsysteme und -logistik der Universität Duisburg-Essen, vorgenommen.

Daten:

Betriebsdaten (OpData), die in einem Intervall von einer Minute im CMS angelegt werden: Betriebsstunden der Krananlagen insgesamt sowie Betriebsstunden der Komponenten: Kranfahrt, Katzfahrt, Drehwerk, Hubwerke, Spreader, Produktivzeit beladen und unbeladen sowie Stillstands- und Wartezeit der Kräne, Wegstrecken Kranfahrt (X-Richtung), Katzfahrt (Y-Richtung), Hubwerk (Z-Richtung) (jeweils in km) und Umdrehungen des Drehwerks, Anzahl der Container- und Trailerumschläge, Anzahl umgeschlagener TEU (Twenty-Foot-Equivalent-Units) und Anzahl umgeschlagener Nutzlasttonnen.

Bewegungsdaten (MoveData), die ereignisbezogen für jede Kranbewegung erfasst werden: Position bei Lastaufnahme und Lastabgabe als XYZ-Koordinate im metrischen System, zurückgelegter Weg der Krananlagen je Umschlag, Umschlagstyp; letzte Position einer vom Kran bewegten Ladungseinheit, ID der Ladungseinheit (Containernummer oder Trailerbezeichnung), Spreaderweite, Drehwinkel. Daten zur Energieeffizienz (MWh): Energieverbrauch jeweils bei Lastaufnahme, von der Lastaufnahme bis zur Lastabgabe als auch die rückgespeiste Energie, Energieverbrauch für die Steuerung und Bewegung und für Nebengewerke (Licht, Klimaanlage etc.), Energierückspeisung aus Senken / Bremsen.

Daten zum Anlagenstatus (StateData), die in einem Intervall von fünf Sekunden erfasst werden. Übersicht zur Verfügbarkeit der Krananlagen (Datentyp boolean): Kran eingeschaltet, produktiv, in Störung, in Parkposition, beladen oder unbeladen; Informationen zu Betriebsmodi: Twistlock- oder Greifzangenbetrieb; Anzahl aktiver Stör- und Warnmeldungen; Daten zu Positionen der Krananlagen: Position Kranfahrt, Katzfahrt, Hubwerk (jeweils in Metern) und Drehwerk (Umdrehungen); aktuelle Last; Abstand zum Nachbarkran bzw. Puffer nach links und rechts; Schiefstellung Katze und Schiefstellung Kranfahrt sowie Höhendifferenz der Hubwerke (jeweils in cm); Ist-Geschwindigkeit (m/min) Kranfahrt, Katzfahrt, Hubwerk, Drehwerk; aktuelle Leistungsaufnahme (kW); aktuelle Windgeschwindigkeit (m/sec) und Windrichtung; aktuelle Außentemperatur.

Daten zu Störungen (FaultsData) der Krananlagen, die ereignisbezogen im CMS hinterlegt werden: von der Störung betroffene Baugruppe und Störmeldungen mit ihren Auswirkungen auf den Betrieb (unterteilt in Klassen: von der Störung mit Ausfall bis hin zur Betriebsinformation ohne Auswirkung), Meldetext der Störmeldung als Klartext, Beginn und Ende der Meldungsursache.

Die Datensätze können im Tabellenformat CSV ausgegeben werden und haben aktuell eine Größe von insgesamt ca. 18 GB (Stand: 08.10.2020).

Verfügbarkeit:

Die Daten können für wissenschaftliche Zwecke im Rahmen des KI-LiveS-Projekts genutzt werden. Für ein Data Repository stehen sie jedoch nicht zur Verfügung.

3.3 Übersicht über weitere im KI-LiveS-Projekt genutzte Datenquellen

Neben institutionsbezogenen Datenquellen, die explizit im Rahmen des KI-LiveS-Projekts erschlossen wurden, werden zur Entwicklung von KI-Tools weitere Datenquellen aus frei zugänglichen Datenbanken³² verwendet. Die nachfolgende Auflistung gewährt einen Überblick über diese Datenquellen und ihre Verwendung im KI-LiveS-Projekt:³³

- MIT-BIH Normal Sinus Rhythm Database³⁴: Verwendung in einem transformativen Data Augmentation Algorithmus zum Training eines Denoising Autoencoders zur Entrauschung von EKG-Daten und als Grundwahrheit;
- MIT-BIH Atrial Fibrillation Database³⁵: Verwendung in einem transformativen Data Augmentation Algorithmus zum Training eines Denoising Autoencoders zur Entrauschung von EKG-Daten und als Grundwahrheit;
- MIT-BIH Noise Stress Test Database³⁶: Verwendung in einem transformativen Data Augmentation Algorithmus zum Training eines Denoising Autoencoders zur Entrauschung von EKG-Daten;
- MIT-BIH Arrhythmia Database³⁷: Training und Test eines CNN-basierten Klassifikators;
- MIT-BIH Compression Test Database³⁸: Testdatensatz für einen Undercomplete (Compression) Autoencoder;
- DUTS Image Dataset³⁹: Training und Test eines Attention-basierten CNNs zur Salient Object Detection.

32) Zu diesen frei zugänglichen Datenbanken gehört vor allem PhysioNet. PhysioNet ist als Internetplattform konzipiert und stellt insbesondere physiologische Daten bereit; vgl. GOLDBERGER/AMARAL/GLASS et al. (2000), S. 215 ff. Daten können auf PhysioNet sowohl frei zugänglich zum Download bereitstehen oder auch erst nach Registrierung für Nutzer downloadbar sein.

33) Die nachfolgend genannten Datenquellen werden im Rahmen des KI-LiveS-Projekts vor allem vom Fachgebiet Eingebettete Systeme der Informatik (UDE) genutzt.

34) Für den Datenursprung vgl. MOODY (1999).

35) Für den Datenursprung vgl. MOODY/MARK (1983), insbesondere S. 228 ff.

36) Für den Datenursprung vgl. MOODY/MULDROW/MARK (1984), S. 381 ff.

37) Für den Datenursprung vgl. MOODY/MARK (2001), S. 45 ff.

38) Für den Datenursprung vgl. MOODY/MARK/GOLDBERGER (1988), S. 167 ff.

39) Für den Datenursprung vgl. WANG/LU/WANG et al. (2017), insbesondere S. 137 ff.

4 Bewertung der Datenquellen

4.1 Datenquellenbezogene Test- und Evaluierungskriterien

Um Datenquellen hinsichtlich ihrer Eignung für Test- und Evaluierungszwecke im Rahmen des KI-LiveS-Projekts zu prüfen, gilt es zunächst, datenquellenbezogene Test- und Evaluierungskriterien zu konzipieren.⁴⁰ Testkriterien dienen dazu, die performative Eignung⁴¹ von Datenquellen in Bezug auf einzelne KI-Tools zu beurteilen. Evaluierungskriterien zielen hingegen darauf ab, Datenquellen hinsichtlich ihrer grundsätzlichen Eignung für die Erfüllung von Forschungs- und Transferzwecken – vor allem im Rahmen des KI-LiveS-Projekts, aber auch „darüber hinaus“ – zu bewerten.

Zunächst werden datenquellenbezogene *Testkriterien* angeführt sowie im Kontext des Teilprojekts „KI-Brainwareentwicklung“⁴² erläutert:

- **Einheitlichkeit:** Die Einheitlichkeit bezieht sich auf die dauerhaft einheitliche Darstellung der in den Datenquellen enthaltenen Daten.⁴³ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Einheitlichkeit auf die dauerhaft einheitliche Darstellung des projektspezifischen Erfahrungswissens in Projektbeschreibung, Projektergebnis und Projektbewertung.
- **Flexibilität:** Die Flexibilität bezieht sich auf die Möglichkeit der separaten Nutzung einzelner Teilmengen von Datenquellen. Im Besonderen zielt dieses Kriterium darauf ab, ob Datenquellen nur in ihrer Gesamtheit zweckmäßig verwendet werden können. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Flexibilität auf die separate Nutzung einzelner Teilmengen von Datenquellen für den Prototyp jCORA als einem exemplarischen KI-Tool.
- **Kompatibilität:** Die Kompatibilität bezieht sich auf die Eignung der Datenquellen, sich in unterschiedliche KI-Soft- und -Hardware integrieren zu lassen. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Kompatibilität auf die Eignung der Datenquellen, in den Prototyp jCORA integrierbar zu sein.

40) Es vermag zunächst erstaunlich anmuten, Test- und Evaluierungskriterien für eine Priorisierung von Datenquellen nachfolgend insbesondere für das Teilprojekt „KI-Brainwareentwicklung“ zu konzipieren, obwohl die Auswertung der Datenquellenerhebung gezeigt hat, dass bis jetzt keine Datenquellen mit projektspezifischem Erfahrungswissen erschlossen werden konnten. Nichtsdestotrotz wird an diesem Vorhaben festgehalten. Dies ergibt sich aus zwei Gründen: Erstens besteht weiterhin die Möglichkeit, im weiteren Projektverlauf Datenquellen mit projektspezifischem Erfahrungswissen zu erschließen; vgl. dazu auch die Ausführungen in Kapitel 6 dieses Projektberichts. Zweitens lässt sich die Konzipierung von Test- und Evaluierungskriterien hinsichtlich der Eignung von Datenquellen für Test- und Evaluierungszwecke im Hinblick auf KI-Tools auch unabhängig vom Kontext des Teilprojekts „KI-Brainwareentwicklung“ nutzen, weil diese Kriterien – erforderlichenfalls in adaptierter Form – auch in anderen Kontexten, wie z. B. anderen Teilprojekten des KI-LiveS-Projekts, als Grundlage für Test- und Evaluierungszwecke dienen können.

41) Unter der performativen Eignung von Datenquellen wird hier verstanden, dass sich die Datenquellen dafür eignen, die Leistungsfähigkeit („Performanz“) von KI-Tools zu beurteilen, die im KI-LiveS-Projekt für den KI-bezogenen Wissenstransfer in die betriebliche Praxis zur Verfügung gestellt werden sollen.

42) Der hier angesprochene Projektkontext erstreckt sich zwar zunächst auf den in der Fußnote 2) priorisierten Gegenstandsbereich des Teilprojekts „KI-Brainwareentwicklung“, kann aber in späteren Überarbeitungen (Auflagen) des Projektberichts Nr. 2 oder auch in Folgepublikationen auf weitere Teilprojekte des KI-LiveS-Projekts ausgedehnt werden.

43) In Anlehnung an ROHWEDER/KASTEN/MALZAHN et al. (2018), S. 26.

- **Konsistenz:** Die Konsistenz bezieht sich hier auf die Widerspruchsfreiheit der Datenquellen.⁴⁴ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Konsistenz auf die Widerspruchsfreiheit des in den Datenquellen enthaltenen projektspezifischen Erfahrungswissens.
- **Machbarkeit:** Die Machbarkeit bezieht sich auf die Machbarkeit der technischen Verwendung der Datenquellen. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Machbarkeit auf die technische Umsetzung des in den Datenquellen gespeicherten Erfahrungswissens im Prototyp jCORA.
- **Persistenz:** Die Persistenz bezieht sich auf die dauerhafte Möglichkeit der Speicherung der Datenquellen. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Persistenz auf die Möglichkeit, dass der Prototyp jCORA dauerhaft auf die Datenquellen mit projektbezogenem Erfahrungswissen zuzugreifen vermag.
- **Sicherheit:** Die Sicherheit bezieht sich auf die Sensibilität der Datenquellen. Im Besonderen sind dahingehend die Anforderungen an die datenschutzkonforme und Unternehmensinteressen schützende Speicherung der Daten zu berücksichtigen. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Sicherheit auf die datenschutzkonforme und Unternehmensinteressen schützende Speicherung der Daten insbesondere im Prototyp jCORA.
- **Strukturierung:** Die Strukturierung bezieht sich auf das Ausmaß der Strukturierung der Datenquellen anhand Unterteilung von strukturierten, semi-strukturierten und unstrukturierten Daten.⁴⁵ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Strukturierung darauf, ob das in den Datenquellen enthaltene Erfahrungswissen in strukturierter, semi-strukturierter oder unstrukturierter Form vorliegt.
- **Vollständigkeit:** Die Vollständigkeit bezieht sich auf das Ausmaß der Vollständigkeit der in den Datenquellen enthaltenen Daten in Bezug auf den Untersuchungsgegenstand.⁴⁶ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Vollständigkeit auf die vollständige Dokumentation des Erfahrungswissens über ein Projekt unter Einhaltung der Strukturierung in Projektbeschreibung, Projektergebnis und Projektbewertung.

Im Folgenden werden die datenquellenbezogenen *Evaluierungskriterien* benannt und im projektspezifischen Kontext des Teilprojekts „KI-Brainwareentwicklung“ erläutert.⁴⁷

- **Eindeutigkeit:** Die Eindeutigkeit bezieht sich auf die Interpretierbarkeit und hierdurch bewirkte Mehrdeutigkeit der Daten von Datenquellen. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Eindeutigkeit auf das Ausmaß der Interpretierbarkeit der in den Datenquellen enthaltenen Daten zu projektspezifischem Erfahrungswissen. Dies kann z. B. eine Angabe im Rahmen einer Projektbewertung sein, dass sich eine Projektdurchführung als „erfolgreich“ erwiesen habe. Solange die jeweils gemeinten Erfolgskriterien nicht präzise offengelegt werden, eröffnet sich ein breiter Interpretationsspielraum, was mit einer „erfolgreichen“ Projektdurchführung konkret gemeint sein könnte.

44) In Anlehnung an HEINRICH/KLIER (2018), S. 58.

45) Vgl. PIRO/GEBAUER (2018), S. 144 f.

46) In Anlehnung an ROHWEDER/KASTEN/MALZAHN et al. (2018), S. 26.

47) Vereinzelt werden im Rahmen der Evaluierungskriterien Kriterien genannt, die zuvor auch schon im Rahmen der Testkriterien genannt worden sind. Eine solche Überschneidung ergibt sich, wenn einem Kriterium sowohl für Test- als auch Evaluierungszwecke Relevanz zugesprochen wird.

- **Einheitlichkeit:** Die Einheitlichkeit bezieht sich auf die dauerhaft einheitliche Darstellung der in den Datenquellen enthaltenen Daten.⁴⁸ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Einheitlichkeit auf die dauerhaft einheitliche Darstellung des projektspezifischen Erfahrungswissens in Projektbeschreibung, Projektergebnis und Projektbewertung.
- **Fehlerfreiheit:** Die Fehlerfreiheit bezieht sich auf die Korrektheit der Angaben innerhalb der Datenquellen.⁴⁹ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Fehlerfreiheit auf die korrekte Dokumentation des projektspezifischen Erfahrungswissens innerhalb der jeweiligen Datenquellen.
- **Konsistenz:** Die Konsistenz bezieht sich auf die Widerspruchsfreiheit der Datenquellen.⁵⁰ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Konsistenz auf die Widerspruchsfreiheit des in den Datenquellen enthaltenen projektspezifischen Erfahrungswissens.
- **Performanz:** Die Performanz bezieht sich auf die Nutzbarkeit der Datenquellen zur weiteren Verwendung durch ein KI-Tool. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ wird hiermit die Möglichkeit bezeichnet, Datenquellen zur Anwendung des Prototyps jCORA zu nutzen.
- **Redundanzfreiheit:** Die Redundanzfreiheit bezieht sich darauf, dass keine Doppelungen innerhalb einer Datenquelle sowie zwischen mehreren Datenquellen stattfinden.⁵¹ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Redundanzfreiheit auf das Ausbleiben von Doppelungen in Bezug auf Datenquellen mit projektspezifischem Erfahrungswissen. Das Evaluierungskriterium der Redundanzfreiheit kann jedoch auch kritisch hinterfragt werden. So kann eine „kontrollierte“ Redundanz durchaus hilfreich sein, um auf benötigte Daten in lokal verfügbaren Datenquellen rascher zugreifen zu können, oder auch, um die Glaubwürdigkeit von Daten auf der Grundlage multipler Datenspeicherung – wie etwa im Rahmen des Blockchain-Konzepts – prüfen zu können.
- **Relevanz:** Die Relevanz bezieht sich auf die Übereinstimmung von Informationsgehalt und Informationsbedarf zwischen den Datenquellen und auf den Forschungs- oder Transferzweck. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Relevanz auf die Übereinstimmung zwischen dem Informationsgehalt der Datenquellen und dem Informationsbedarf, der zur Evaluierung des Prototyps jCORA besteht.⁵²
- **Repräsentativität:** Die Repräsentativität bezieht sich auf den möglichen Rückschluss von Erkenntnissen, die aus einer Stichprobe gewonnen wurden, auf die zugrunde liegende (aber weitgehend unbekannt) Grundgesamtheit. Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Repräsentativität auf die Möglichkeit des Rückschlusses des in

48) In Anlehnung an ROHWEDER/KASTEN/MALZAHN et al. (2018), S. 26.

49) In Anlehnung an STORCH/LAUE/GRUHN (2013), S. 2.

50) Vgl. HEINRICH/KLIER (2018), S. 58.

51) In Anlehnung an WOLF (2018), S. 243.

52) Der Informationsbedarf im Rahmen des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich zunächst, wie schon in Kapitel 2.1 skizziert wurde, auf projektbezogenes Erfahrungswissen. Dieses projektbezogene Erfahrungswissen wird für das Teilprojekt „KI-Brainwareentwicklung“ aus zwei Gründen benötigt: Einerseits gilt es, aus bestehendem projektbezogenem Erfahrungswissen sprachliche Ausdrucksmittel aus der betrieblichen Praxis zu extrahieren, um diese sprachlichen Ausdrucksmittel in die im weiteren Projektverlauf zu entwickelnde Projektmanagement-Ontologie zu integrieren. Andererseits gilt es, bestehendes projektbezogenes Erfahrungswissen in das ontologiegestützte Case-based-Reasoning-System jCORA zu integrieren, um zu evaluieren, ob das prototypische KI-Tool jCORA tatsächlich die „intelligente“ Wiederverwendung von Erfahrungswissen ermöglicht.

den Datenquellen enthaltenen projektspezifischem Erfahrungswissens auf eine Grundgesamtheit⁵³ von Projekten, die seitens der betrieblichen Praxis für den Einsatz eines ontologiegestützten Case-based-Reasoning-Systems wie jCORA für relevant erachtet werden.

- **Verständlichkeit:** Die Verständlichkeit bezieht sich auf das Ausmaß der intuitiven Verständlichkeit durch den Nutzer.⁵⁴ Im Kontext des Teilprojekts KI-Brainwareentwicklung bezieht sich die Verständlichkeit auf das Ausmaß der intuitiven Verständlichkeit des in den Datenquellen enthaltenen projektspezifischen Erfahrungswissens.
- **Vollständigkeit:** Die Vollständigkeit bezieht sich auf das Ausmaß der in den Datenquellen enthaltenen Daten in Bezug auf den Untersuchungsgegenstand.⁵⁵ Im Kontext des Teilprojekts „KI-Brainwareentwicklung“ bezieht sich die Vollständigkeit auf die vollständige Dokumentation aller betriebswirtschaftlich relevanten Aspekte eines Projekts unter Einhaltung der Strukturierung in Projektbeschreibung, Projektergebnis und Projektbewertung.

4.2 Priorisierung der Nutzung von Datenquellen

4.2.1 Auswahl einer Methode zur Priorisierung von potenziellen Datenquellen

Um aus der Vielzahl möglicher Datenquellen die zweckmäßigsten für das KI-LiveS-Projekt oder einzelne Teilprojekte auswählen zu können, erfolgt eine kriterienorientierte Priorisierung der potenziellen Datenquellen. Für eine solche kriterienorientierte Priorisierung empfiehlt sich der Einsatz multi-kriterieller Bewertungsmethoden. Hierzu stehen verschiedene multikriterielle Bewertungsmethoden zur Verfügung, die nachfolgend genannt und kurz skizziert werden:⁵⁶

- Analytic Hierarchical Process (AHP)
- Checklistenverfahren
- Goal Programming
- Nutzwertanalyse (NWA) oder Scoring-Methode
- Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)
- Vektoroptimierung

Der von SAATY entwickelte Analytic Hierarchy Process (AHP) stellt eine Bewertungsmethode zur Lösung multikriterieller Bewertungsprobleme dar.⁵⁷ Beim AHP erfolgt zunächst eine Dekomposition des jeweils vorliegenden Bewertungsproblems in eine Hierarchie von Bewertungskriterien und – auf der tiefsten Hierarchiestufe – von zu bewertenden Alternativen.⁵⁸ Auf der Grundlage dieser Hierarchie erfolgt eine paarweise Beurteilung von jeweils zwei Elementen einer Hierarchieebene in Bezug

53) Für das Festlegen der Grundgesamtheit lassen sich mehrere Varianten in Betracht ziehen. In intuitiver Hinsicht liegen am nächsten die Gesamtheit an Projekten eines Unternehmens oder einer Branche. Die Definition der Grundgesamtheit ist hier stets im Einzelfall der Anwendung eines KI-Tools vorzunehmen.

54) In Anlehnung an ROHWEDER/KASTEN/MALZAHN et al. (2018), S. 27.

55) In Anlehnung an ROHWEDER/KASTEN/MALZAHN et al. (2018), S. 26.

56) Die Anzahl denkmöglicher multikriterieller Bewertungsverfahren übersteigt die in diesem Projektbericht vorgestellten multikriteriellen Bewertungsmethoden um ein Vielfaches. Daher wurde zur Wahrung der Übersichtlichkeit eine „erfahrungsbasierte“ Vorauswahl der vorgestellten multikriteriellen Bewertungsmethoden durch die Verfasser des Projektberichts getroffen.

57) Vgl. z. B. SAATY (1986), S. 841 ff.; SAATY (1990), S. 10 ff.; SAATY (1994), insbesondere S. 21 ff.; SAATY (2000), insbesondere S. 1 ff. u. S. 45 ff.; SAATY (2008), insbesondere S. 14 ff.; SAATY/VARGAS (1997); SAATY/VARGAS (1998), insbesondere S. 493 ff.

58) Vgl. SAATY (1994), S. 22 f.

auf ein unmittelbar übergeordnetes Element der Hierarchie.⁵⁹ Die Paarvergleichsurteile werden in sogenannte Evaluationsmatrizen überführt, auf deren Grundlage Gewichtungen (lokale Prioritäten) für einzelne Bewertungskriterien und für Alternativen in Bezug auf die jeweils unmittelbar übergeordneten Bewertungskriterien ermittelt werden.⁶⁰ Lokale Prioritäten von Bewertungskriterien werden für die Ermittlung ihrer Bedeutung in Bezug auf die Gesamthierarchie des zugrunde liegenden Bewertungsproblems mit allen lokalen Prioritäten der ihnen auf einem Pfad übergeordneten Bewertungskriterien zu einer globalen Priorität multiplikativ zusammengefasst. Diese globale Priorität eines Bewertungskriteriums wird schließlich mit der lokalen Priorität einer Alternative zu einem Teilnutzen multipliziert.⁶¹ Die Summe aller Teilnutzen einer Alternative hinsichtlich aller Bewertungskriterien auf der untersten Ebene der Kriterienhierarchie stellt schließlich die globale Priorität oder – synonym verstanden – den Gesamtnutzen einer Alternative dar.

Die Checklisten-Methode gilt als eine der einfachsten Methoden zur Lösung multikriterieller Bewertungsprobleme. Bei der Checklisten-Methode werden die als relevant erachteten Bewertungskriterien übersichtlich aufgelistet und ihre Erfüllungen durch die Alternativen weitgehend intuitiv beurteilt. Um eine Auswahl einer Alternative zu treffen, können verschiedene Heuristiken angewendet werden. Beispielsweise kann die Alternative mit der höchsten Anzahl an erfüllten Bewertungskriterien ausgewählt werden. Ebenso denkbar ist die Auswahl aller derjenigen Alternativen, die eine zu spezifizierende Mindestanzahl an Bewertungskriterien erfüllt.

Das Goal Programming ist ein mathematisches Optimierungsverfahren, welches auf die Arbeit von CHARNES/COOPER⁶² zurückzuführen ist. Im Zuge des Goal Programmings werden zunächst für eine Vielzahl von Zielen – analog Bewertungskriterien – anzustrebende Soll-Werte festgelegt.⁶³ Für diese Ziele werden folgend die Ist-Werte, welche durch die jeweiligen Alternativen zu erwarten sind, ermittelt. Alsdann werden die Abweichungen der Ist- von den Soll-Werten ermittelt. Das Ziel ist die Minimierung der alternativenspezifischen Gesamtabweichung zwischen gewichteten Soll- und Ist-Werten in Bezug auf alle Bewertungskriterien.

Die Nutzwertanalyse stellt eine multikriterielle Bewertungsmethode dar, bei der eine Menge von Alternativen entsprechend den (Bewertungs-)Präferenzen des Entscheidungsträger systematisch geordnet wird.⁶⁴ Die Ordnung der Alternativen erfolgt anhand von sogenannten Nutzwerten.⁶⁵ Die Nutzwerte beruhen auf einer ganzheitlichen Alternativenbewertung auf der Grundlage aller Bewertungskriterien des Entscheidungsträgers. Pro Alternative wird für jedes Bewertungskriterium ein Teilnutzwert ermittelt.⁶⁶ Durch Aufsummieren der gewichteten Teilnutzenwerte ergibt sich der Gesamtnutzenwert einer Alternative.

Die Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) ist eine Methode zur Lösung multikriterieller Bewertungsprobleme, insbesondere zur ökonomischen Effizienzanalyse (Bewertung der Effizienz von betriebs- oder volkswirtschaftlichen Analyseeinheiten, wie z. B. Unternehmen bzw. Gesundheitssystemen). Sie geht vor allem auf Arbeiten von HWANG/YOON⁶⁷ zurück. Grundlage für die Anwendung der Bewertungsmethode TOPSIS ist die Definition einer jeweils

59) Vgl. SAATY (1994), S. 25 ff.

60) Vgl. RIEDL (2006), S. 106 f. Das Zitat ist auch für den nachfolgenden Satz gültig.

61) Vgl. RIEDL (2006), S. 110. Das Zitat ist auch für den nachfolgenden Satz gültig

62) Vgl. CHARNES/COOPER (1961), S. 215 ff.

63) Vgl. ZELEWSKI/PETERS (2005), S. 1208. Das Zitat ist bis zum Absatzende gültig.

64) Vgl. SCHNEEWEIß (1990), S. 13; ZANGEMEISTER (1976), S. 45.

65) Vgl. ZANGEMEISTER (1976), S. 45. Das Zitat ist auch für den nachfolgenden Satz gültig.

66) Vgl. ZANGEMEISTER (1976), S. 71 ff.; siehe auch RIEDL (2006), S. 111 ff. Das Zitat ist auch für den nachfolgenden Satz gültig.

67) Vgl. HWANG/YOON (1981). S. 128 ff.

idealisierten besten und schlechtesten Alternative.⁶⁸ Die Vorteilhaftigkeit einer realen Alternative wird durch einen möglichst geringen Abstand zur besten sowie einen möglichst großen Abstand zur schlechtesten idealisierten Alternative gemessen.⁶⁹

Bei der Vektoroptimierung⁷⁰ wird aus einer Menge von zu bewertenden Alternativen mindestens eine Alternative ausgewählt, die sich im Hinblick auf die zugrundeliegenden Bewertungskriterien als nicht-dominierte Alternative herausstellt.⁷¹ Eine Alternative gilt als nicht-dominiert, wenn es keine andere Alternative gibt, die hinsichtlich mindestens eines Bewertungskriteriums als besser und hinsichtlich aller anderen Bewertungskriterien als mindestens gleich gut beurteilt wird.

Aus der Vielzahl der vorgestellten skizzierten multikriteriellen Bewertungsmethoden wird für die Priorisierung der Datenquellen nachfolgend die Nutzwertanalyse verwendet. Diese Methodenauswahl beruht auf der intuitiv leicht verständlichen und zeitsparenden Anwendungsmöglichkeit der Nutzwertanalyse. Da die Nutzwertanalyse von den skizzierten multikriteriellen Bewertungsmethoden als jene anzusehen ist, die am bekanntesten und verbreitetsten einzuschätzen ist, erlaubt ihre Verwendung das Entwickeln eines Anwendungsschemas, das über die Priorisierung von Datenquellen im spezifischen Projektkontext des KI-LiveS-Projekts hinaus von Dritten auch in verwandten Anwendungsgebieten eingesetzt werden kann.

68) Vgl. WANG/LEE (2007), S. 1763.

69) Vgl. GHOSH (2011), S. 64.

70) In der einschlägigen Fachliteratur findet sich oftmals auch der Terminus Effizienzanalyse.

71) Vgl. GÖPFERT/NEHSE (1990), S. 17 ff.; siehe auch DINKELBACH (1969), S. 63 ff.

4.2.2 Anwendungsschema für die Priorisierungsmethode der Nutzwertanalyse

Nachfolgend wird in tabellarischer Form ein Anwendungsschema für die Priorisierung von Datenquellen für Testzwecke einerseits und Evaluierungszwecke andererseits vorgestellt. Auf eine ursprünglich intendierte Priorisierung der erschlossenen Datenquellen, insbesondere für das Teilprojekt „KI-Brainwareentwicklung“, wird an dieser Stelle verzichtet, weil – wie bereits in Kapitel 3.1 dargestellt wurde – keine Datenquellen mit projektspezifischem Wissen erschlossen werden konnten. Vielmehr dient die nachfolgende tabellarische Darstellung des Anwendungsschemas der Nutzwertanalyse lediglich dem Aufzeigen des intendierten Vorgehens zur Priorisierung von Datenquellen. Dazu wird in den Tabellen 2 und 3 jeweils ein Anwendungsschema für die Nutzwertanalyse vorgestellt, das sich auf die Priorisierung von Datenquellen für entweder Test- oder aber Evaluationszwecke bezieht.

Testkriterien	Gewichtung	Alternative 1		Alternative 2		Alternative 3		Alternative 4	
		Bewertung	Teilnutzwert	Bewertung	Teilnutzwert	Bewertung	Teilnutzwert	Bewertung	Teilnutzwert
Einheitlichkeit									
Flexibilität									
Kompatibilität									
Konsistenz									
Machbarkeit									
Persistenz									
Sicherheit									
Strukturierung									
Vollständigkeit									
Summe									

Tabelle 2: Anwendungsschema zur Nutzwertanalyse für Testzwecke

Für die Bestimmung der Gewichte der Testkriterien wird eine für die Nutzwertanalyse typische fünfstufige Skala mit den Punktwerten „1“ bis „5“ bevorzugt, wobei „1“ für ein sehr geringes und „5“ für ein sehr hohes Gewicht steht. Für die Bewertung der Alternativen wird analog zu der Gewichtung der Testkriterien ebenso eine fünfstufige Skala mit den Punktwerten „1“ bis „5“ angewendet.

Evaluierungskriterien	Gewichtung	Alternative 1		Alternative 2		Alternative 3		Alternative 4	
		Bewertung	Teilnutzwert	Bewertung	Teilnutzwert	Bewertung	Teilnutzwert	Bewertung	Teilnutzwert
Aufbereitungsgrad									
Eindeutigkeit									
Einheitlichkeit									
Fehlerfreiheit									
Konsistenz									
Performanz									
Redundanzfreiheit									
Relevanz									
Verständlichkeit									
Vollständigkeit									
Summe									

Tabelle 3: Anwendungsschema zur Nutzwertanalyse für Evaluierungszwecke

Für die Anwendung der Nutzwertanalyse zur Priorisierung von Datenquellen für Evaluierungszwecke ist das zuvor vorgestellte Anwendungsschema, das sich auf die Nutzwertanalyse zur Priorisierung von Datenquellen für Testzwecke bezog, analog anzuwenden.

5 Evaluation potenzieller Datenquellen

Hinsichtlich der Evaluation potenzieller Datenquellen ist zu konstatieren, dass sowohl die Quantität als auch die Qualität identifizierter Datenquellen weit hinter den Erwartungen des Projektkonsortiums des KI-LiveS-Projekts zurückgeblieben ist.

Zunächst wird das Teilprojekt „KI-Brainwareentwicklung“ betrachtet. Aus dieser Perspektive ist es leider nicht gelungen, Datenquellen mit projektspezifischem Erfahrungswissen zu identifizieren, welche die oben angesprochenen Test- oder Evaluierungskriterien auch nur „angenähert“ erfüllen, und anschließend „teilprojektwirksam“ zu erschließen.⁷² Daher erfolgt in diesem Projektbericht auch keine konkrete Anwendung des Schemas, das in Kapitel 4.2.2 für die Anwendung einer Nutzwertanalyse auf die Auswahl von „geeigneten“ Datenquellen vorgestellt wurde. Weil die Basis der identifizierten Datenquellen letztlich einer „leeren Menge“ entsprach, entfiel die Aufgabe, aus zunächst identifizierten potenziellen Datenquellen (mindestens) eine geeignete Datenquelle mittels einer multikriteriellen Bewertungsmethode auszuwählen.

Die Gründe für die mangelnde Identifizierung von Datenquellen mit projektspezifischem Erfahrungswissen sind vielfältiger Natur. Zunächst ist selbstkritisch anzumerken, dass sich im Zuge der Corona-Pandemie die Zusammenarbeit mit Praxispartnern als äußerst schwierig erwies. Dies ist einerseits darauf zurückzuführen, dass eine intensive Zusammenarbeit mit den Praxispartnern vor Ort zum Verdeutlichen⁷³ der Wichtigkeit der Bereitstellung von Datenquellen für die Entwicklung eines KI-Tools zur „intelligenten“ Wiederverwendung von Erfahrungswissen zu weiten Teilen, anders als ursprünglich geplant, nicht möglich war. Andererseits lagen im Zuge der Corona-Pandemie die Prioritäten von Unternehmen verständlicherweise zunächst auf der Restrukturierung vorhandener Geschäftsprozesse zur Sicherung ihrer Wettbewerbsfähigkeit. Daher standen unternehmensseitige Kapazitäten für die Teilnahme an einem universitären Forschungs- und Transferprojekt oftmals nicht zur Verfügung.

Unabhängig von der Corona-Pandemie war der allgemeine Tenor seitens der Praxispartner, dass die Datenquellen⁷⁴, die für die Nutzung eines ontologiegestützten Case-based-Reasoning-Systems wünschenswert sind, zumeist hochsensible Unternehmens- und Kundendaten beinhalten. Aus nachvollziehbaren und schützenswerten Unternehmensinteressen wäre es unrealistisch anzunehmen, diese Datenquellen im Rahmen eines Data Repository allgemein oder auch nur nutzergruppenspezifisch (wie z. B. innerhalb der Nutzergruppe der assoziierten Praxispartner des KI-LiveS-Projekts) nutzen zu können. Eine solche Haltung entspricht der oftmals zu beobachtenden Haltung, Daten als ein „Geschäftsgeheimnis“ zu betrachten und diese vor Einblicken Dritter zu bewahren. Dies vermag in

72) Als Randnotiz sei an dieser Stelle vermerkt, dass die fehlende Bereitschaft zur Verfügungstellung von Datenquellen mit projektspezifischem Erfahrungswissen für das Projektteam des Teilprojekts „KI-Brainwareentwicklung“ nicht gänzlich überraschend ist. Schon im Zuge des Projektberichts Nr. 1 des KI-LiveS-Projekts, dessen Fokus auf einer Anforderungsanalyse für ein KI-Tool zur „intelligenten“ Wiederverwendung von Erfahrungswissen lag, jedoch auch am Rande die Bereitschaft zur Bereitstellung von Datenquellen mit projektspezifischem Erfahrungswissen erfragte, wurde die Bereitschaft zu jener Datenquellenbereitstellung als sehr gering oder unklar eingestuft. Dennoch hatten Praxispartner vereinzelt ihre Bereitschaft signalisiert, zumindest eine kleine Anzahl an Datenquellen mit projektspezifischem Erfahrungswissen zur Verfügung zu stellen; vgl. dazu die Ausführungen in SCHAGEN/ZELEWSKI/HEEB (2020), S. 57 f.

73) Aus der Sicht des Projektteams des Teilprojekts „KI-Brainwareentwicklung“ ist insbesondere die Kooperation mit den Praxispartnern „vor Ort“ von entscheidender Bedeutung für die Weiterentwicklung des Prototyps jCORA für ein ontologiegestütztes Case-based-Reasoning-System. Dieser entscheidende Charakter begründet sich dadurch, dass nur durch eine prototypische Implementierung in Unternehmen die Eignung des Prototyps jCORA für die betriebliche Nutzung durch erfahrene Projektmitarbeiter(innen) evaluiert werden kann. Hinsichtlich des Erschließens potenzieller Datenquellen kann zudem durch eine prototypische Implementierung in Unternehmen die Notwendigkeit der Bereitstellung von Datenquellen mit projektspezifischem Erfahrungswissen verdeutlicht werden und so möglicherweise die Bereitschaft zur Bereitstellung von Datenquellen mit projektspezifischem Erfahrungswissen seitens der Praxispartner gefördert werden.

74) Vgl. für solche Anforderungen an Datenquellen die in Kapitel 2.1 dargestellte Struktur der Datenquellen.

Zeiten, in denen beispielsweise „Big Data“ und „Shared Economy“ zu den (vermeintlichen) Mega-Trends gehören und „in aller Munde“ sind, überraschen, ist aber seitens des Projektteams des Teilprojekts „KI-Brainwareentwicklung“ zunächst so hinzunehmen. Dennoch ist zu konstatieren, dass der Projekterfolg des Teilprojekts „KI-Brainwareentwicklung“ wegen der fehlenden Bereitstellung von Datenquellen mit projektspezifischem Erfahrungswissen beeinträchtigt sein könnte. Diese Gefährdung beruht darauf, dass eine realitätsnahe Erprobung des Prototyps jCORA für ein KI-Tool im Hinblick auf die Eignung, in der betrieblichen Praxis eingesetzt werden zu können, ohne projektbezogenes Erfahrungswissen aus der betrieblichen Praxis nur in erheblich eingeschränkter Weise geleistet werden kann. Allerdings haben die Mitglieder des Teilprojekts „KI-Brainwareentwicklung“ erste Kontakte zu Unternehmen (Atos Information Technology GmbH, GFOS mbH, IT.NRW und DMT GmbH & Co. KG) geknüpft, die darauf hoffen lassen, in den nächsten Monaten der Projektlaufzeit auch einen selektiven (aus Unternehmenssicht stark eingeschränkten) Zugriff auf unternehmensspezifische Datenquellen mit projektspezifischem Erfahrungswissen zu erlangen.

Hinsichtlich weiterer Teilprojekte des KI-LiveS-Projekts lassen sich die erschlossenen Datenquellen sowohl hinsichtlich ihrer Qualität als auch hinsichtlich ihrer Quantität als „ausbaufähig“ beurteilen. Jedoch bieten die vielfältigen identifizierten Datenquellen, die vor allem in Kapitel 3.2 (aber auch in Kapitel 3.3) angesprochen wurden, zahlreiche und vielversprechende Ansatzpunkte, um das Potenzial von KI-Tools zu testen und hinsichtlich ihrer Eignung für Herausforderungen der betrieblichen Praxis sowie in Bezug auf die Forschungs- und Transferzwecke des KI-LiveS-Projekts zu evaluieren. Zum gegenwärtigen Zeitpunkt werden zur weiteren Konkretisierung möglicher Einsatzpotenziale bilaterale Gespräche zwischen den betroffenen Universitätspartnern des Projektkonsortiums des KI-LiveS-Projekts und den potenziellen „Datenbereitstellern“, vor allem den assoziierten Praxispartnern des KI-LiveS-Projekts, geführt. Diese Gespräche stimmen optimistisch, im weiteren Projektverlauf zusätzliche Datenquellen für Test- und Evaluierungszwecke nicht nur identifizieren, sondern auch hinsichtlich ihrer operativen Nutzung erschließen zu können.

6 Ausblick auf weitere Datenquellen für ein „Data Repository“ im Verlauf des KI-LiveS-Projekts

Für den zukünftigen Verlauf des KI-LiveS-Projekts ist beabsichtigt, weitere Datenquellen für das Data Repository des KI-Labors zu erschließen.

Für das Teilprojekt „KI-Brainwareentwicklung“ werden weitere Versuche unternommen werden, Datenquellen für projektspezifisches Erfahrungswissen zu erschließen. Als potenzielle Möglichkeiten, weiteres projektspezifisches Erfahrungswissen zu erschließen, bieten sich im Besonderen Akquiseversuche sowohl wiederholt bei den aktuellen assoziierten Praxispartnern als auch bei zukünftig möglichen assoziierten Praxispartnern (wie z. B. der DMT GmbH & Co. KG) des Teilprojekts „KI-Brainwareentwicklung“ an; vgl. dazu die Ausführungen in Kapitel 5. Ob solche Akquiseversuche bis zum Projektende ertragreich sein werden, lässt sich zum gegenwärtigen Zeitpunkt nicht seriös abschätzen.

In Anbetracht der in Kapitel 5 identifizierten Ursachen für die bisherigen Schwierigkeiten hinsichtlich des Erschließens von Datenquellen für projektspezifisches Erfahrungswissen lässt sich im Hinblick auf die Corona-Pandemie zunächst Folgendes konstatieren: Zum gegenwärtigen Zeitpunkt ist nicht absehbar, ob bis zum Projektende eine „Normalisierung“ der Beziehungen zu den Praxispartnern in Bezug auf die Corona-Pandemie möglich sein wird. Jedoch besteht die Perspektive, dass sich aufgrund der sowohl auf Praxis- als auch auf Universitätsseite mittlerweile weitaus stärker etablierten digitalen Kommunikationskanäle (wie z. B. Telekonferenzen im Internet) eine verstärkte Kommunikation mit den Praxispartnern realisieren lässt. Im Zuge einer solchen vertieften Kommunikation könnte im Hinblick auf das Erschließen von Datenquellen für projektspezifisches Erfahrungswissen eine neue Dynamik entstehen. Erste Gesprächstermine mit ausgewählten Praxispartnern u. a. im Hinblick auf die Bereitstellung von Datenquellen mit projektspezifischem Erfahrungswissen sind dazu seitens des Instituts für Produktion und Industrielles Informationsmanagement schon vereinbart. Sie betreffen insbesondere die GFOS mbH, mit der eine vertiefte Kooperation im Teilprojekt „KI-Brainwareentwicklung“ ab Februar 2021 vereinbart wurde.

Ob im Zuge weiterer Gespräche mit den Praxispartnern Vorbehalte im Hinblick auf die in Kapitel 5 skizzierte Betrachtung von Daten als „Geschäftsgeheimnisse“ ausgeräumt werden können, wird seitens des Projektteams des Teilprojekts „KI-Brainwareentwicklung“ zum gegenwärtigen Zeitpunkt erheblich bezweifelt. Das Narrativ der „Wahrung von Geschäftsgeheimnissen“ ist in der betrieblichen Praxis gegenwärtig „alternativlos“ dominant. Um eine Abwendung von diesem Narrativ der „Geheimniswahrung“ zu erreichen, sind vermutlich erhebliche organisatorische Veränderungen in Unternehmen erforderlich.

Um den Vorbehalten auf Unternehmensseite im Hinblick auf eine Mitarbeit an der Weiterentwicklung eines KI-Tools zur „intelligenten“ Wiederverwendung von projektspezifischem Erfahrungswissen vorzubeugen, bietet es sich an, auf immer populärer werdende „kooperative“ Organisationsstrukturen oder Unternehmenskulturen zurückzugreifen. Dazu gehören vor allem „virtuelle Unternehmen“ sowie „Communities of Practice“.

Virtuelle Unternehmen zeichnen sich durch den Zusammenschluss mehrerer rechtlich unabhängiger Unternehmen aus, die für die „Produktion“ einer bestimmten Leistung eine vertrauensbasierte Kooperation eingehen und aus der Kundensicht wie nur ein („virtuelles“) Unternehmen agieren.⁷⁵ Eine solche unternehmensübergreifende Organisationsstruktur könnte – in entsprechend adaptierter Weise – auch auf die Zusammenarbeit mehrerer Praxispartner hinsichtlich der Weiterentwicklung eines KI-Tools zur „intelligenten“ Wiederverwendung von projektspezifischem Erfahrungswissen übertragen werden und den ökonomischen Nutzen einer solcher Zusammenarbeit unterstreichen.

75) Vgl. BORCHARDT (2006), S. 17.

Eine ähnliche projektspezifische Adaption ließe sich auch aus dem Konzept einer Community of Practice ableiten. Communities of Practice sind ein Zusammenschluss verschiedenener Akteure, die in Bezug auf eine spezielle Domäne der betrieblichen Praxis den Austausch von Lernprozessen und so eine Wissenserweiterung anstreben.⁷⁶ Charakteristisch für Communities of Practice sind ein gemeinsam geteiltes Verständnis über das Ziel der eigenen Community, über das gegenseitige Einbringen eines gewissen Ausmaßes an Leistungsbereitschaft zur Etablierung von Normen und Standards innerhalb der Community sowie über eine gemeinsame Produktion von Wissensressourcen als Ergebnis einer solchen Community. Den Gedanken aufgreifend, könnte es auch für Unternehmen von Interesse sein, eine Art Community of Practice für den Bereich der „intelligenten“ Wiederverwendung von Erfahrungswissen zu bilden und auf Basis gemeinsam geteilter Normen und Standards sowie einer vertrauensbasierten Zusammenarbeit an der (Weiter-)Entwicklung eines KI-Tools zur „intelligenten“ Wiederverwendung von Erfahrungswissen mitzuwirken.

Resümierend besteht für das Teilprojekt „KI-Brainwareentwicklung“ aufgrund der zuvor skizzierten Argumente durchaus eine realistische Erwartung, bis zum Projektende Datenquellen mit projektspezifischem Erfahrungswissen zu erschließen. Ob und in welchem Umfang dies gelingen wird, kann jedoch gegenwärtig nicht abschließend beantwortet werden.

Für die weiteren Teilprojekte des KI-LiveS-Projekts sind zum gegenwärtigen Zeitpunkt keine weiteren Akquiseversuche zur Erschließung potenzieller Datenquellen für Test- und Evaluierungszwecke von KI-Tools bei anderen Praxispartnern bekannt. Dies kann sich in den kommenden Monaten der Projektlaufzeit jedoch noch ändern.

76) Vgl. WENGER (2000), S. 229. Das Zitat ist auch für den nachfolgenden Satz gültig.

7 Möglichkeiten für ein Data Repository im KI-LiveS-Projekt

Für den Aufbau eines Data Repository zur Archivierung und Wiederverwendung von Daten im Rahmen des KI-LiveS-Projekts gilt es, eine Vielzahl von Entscheidungen zu treffen. Solche Entscheidungen beziehen sich beispielsweise auf die Art des Repository, auf Zugangsmöglichkeiten unter Beachtung datenschutzrechtlicher Bestimmungen, auf die Einbettung des Data Repository in eine zuverlässige IT-Umgebung sowie auf die Verwaltungs- und Nutzungsmöglichkeiten des Data Repository insbesondere auch nach dem Ende des KI-LiveS-Projekts. Solche Entscheidungen können im Rahmen dieses Projektberichts nicht abschließend erfolgen. Vielmehr ist intendiert, nachfolgend einige Möglichkeiten aufzuzeigen, die im Rahmen einzelner, ausgewählter⁷⁷ Entscheidungen für ein Data Repository diskutiert werden sollten.

Zunächst ist die Art des Data Repository zu diskutieren. Grundsätzlich lässt sich eine Vielzahl von Data Repositories unterscheiden. Zu den gängigsten Arten gehören „institutional repositories“ im Sinne von Publikationsservern, „trusted repositories“ für die geschützte Langzeitarchivierung von Daten sowie „open access repositories“, die für frei zugängliche Daten konzipiert sind.⁷⁸ Im Sinne des KI-LiveS-Projekts sind vor allem die beiden letztgenannten Repository-Arten von Bedeutung. Welcher der genannten Arten eines Repositories präferiert wird, hängt insbesondere von den Zugangsmöglichkeiten ab, die für das Data Repository des KI-LiveS-Projekts angestrebt werden. Eine solche Entscheidung ist von zwei gegenläufigen Interessensgruppen geprägt. Einerseits besteht seitens der Universitätseite oftmals ein großes Interesse an frei zugänglichen Forschungs- und Transferergebnissen im Sinne des „Open Access“. Die Universität Duisburg-Essen als übergeordnete Organisation der Mehrzahl der Universitätspartner des Projektkonsortiums des KI-LiveS-Projekts spricht sich beispielsweise für einen offenen Umgang und freien Zugang zu wissenschaftlichen Ergebnissen und Forschungsdaten aus.⁷⁹ Andererseits sind insbesondere die im KI-LiveS-Projekt benötigten Unternehmensdaten von besonders schützenswertem Interesse und seitens der Unternehmen in der Regel für einen solchen freien Zugang nicht vorgesehen. Neben diesen schützenswerten Unternehmensinteressen unterliegt die Langzeitarchivierung von Daten zudem datenschutzrechtlichen Bestimmungen, die Zugriffsbeschränkungen auf ein solches Data Repository erfordern können.⁸⁰

Neben der Art des Data Repository ist zu entscheiden, ob das Data Repository des KI-LiveS-Projekts als ein neues, eigenständiges Data Repository zu verstehen ist, oder ob sich das Projektkonsortium dazu entscheidet, sich einem bestehendem Data Repository anzuschließen. Um ein geeignetes⁸¹, bestehendes Data Repository zu identifizieren, kann eine Online-Datenbank zum Auffinden von Data Repositories verwendet werden.⁸² Die Universität Duisburg-Essen bietet mittels ihres Informations- und Mediendienstes ZIM zumindest im Bereich der Publikationsrepositorien eine Möglichkeit, an einem bestehendem Data Repository zu partizipieren. Darüber hinaus werden Universitätsangehörige

77) Nachfolgend wird lediglich ein Teil der potenziell zu treffenden Entscheidungen für den Aufbau des Data Repository des KI-LiveS-Projekts angeschnitten.

78) Vgl. ASCHENBRENNER/NEUROTH (2011), S. 102 f.

79) Vgl. Universität Duisburg-Essen (2020).

80) Vgl. KLUMP (2011), S. 117 f.

81) Um die Eignung bestehender Data Repositories beurteilen zu können, müssten zunächst entsprechende Eignungskriterien entwickelt werden. Auf die Entwicklung solcher Eignungskriterien wird an dieser Stelle verzichtet, weil zunächst eine Entscheidung im Hinblick auf die vorangestellte Frage der Eigenständigkeit des Data Repository zu treffen ist.

82) Als eine der größten Online-Datenbanken zur Suche von Data Repositories gilt „re3data“. Die Datenbank „re3data“ wurde von der Deutschen Forschungsgemeinschaft (DFG) gegründet. Sie bietet die Suche von Data Repositories in vielen verschiedenen, akademischen Fachrichtungen an; vgl. re3data.org (2021).

gen verschiedene Möglichkeiten zur Speicherung ihrer Forschungsdaten offeriert. Zu diesen Möglichkeiten der Datenspeicherung gehören insbesondere Fileserver, Bandspeicher, Objektspeicher und Datenbanken.⁸³

Weitere Entscheidungen gilt es hinsichtlich der zugrundeliegenden Data-Repository-Software zu treffen. Zu den gängigsten Data-Repository-Softwares gehören EPrints⁸⁴, Dspace⁸⁵, Fedora⁸⁶ und iRODS^{87, 88}. Im Hinblick auf die Nutzung für ein Data Repository wird insbesondere diesen Softwares eine grundsätzliche Eignung zugesprochen.⁸⁹

Wesentlich bei dem Aufbau eines Data Repository ist die Einhaltung von datenschutzrechtlichen Bestimmungen. In Deutschland beruhen datenschutzrechtliche Bestimmungen auf der Datenschutz-Grundverordnung (DSGVO) der EU und ihrer Umsetzung in deutsches Recht, aus dem Bundesdatenschutzgesetz und spezifischen Landesdatenschutzgesetzen; sie richten sich in erster Linie auf den Schutz personenbezogener Daten.⁹⁰ Neben der Einhaltung datenschutzrechtlicher Bestimmungen gilt es darüber hinaus, die Datensicherheit für ein Data Repository zu gewährleisten. Die Datensicherheit ist hinsichtlich der inhaltlichen Ausgestaltung weiter gefasst als der Datenschutz und meint den Schutz von Daten jeglicher Art.⁹¹ Da oftmals ein Spannungsfeld zwischen dem Einhalten datenschutzrechtlicher Bestimmungen und der Nutzung von Daten für wissenschaftliche Zwecke besteht, gibt es für Forschungszwecke datenschutzrechtliche Sonderbestimmungen, die weiterhin zu beachten sind.⁹²

Neben solchen Entscheidungen hinsichtlich des Aufbaus eines Data Repository sollen gängige Standards im Umgang mit Forschungsdaten eingehalten werden. Zu solchen Standards gehören im Besonderen die Einhaltung der sogenannten FAIR-Prinzipien⁹³. Sie umfassen vor allem vier übergeordnete Prinzipien:⁹⁴

- Findable: Die Daten sollen für Dritte auffindbar sein. Um diesem Prinzip gerecht zu werden, sollen die Daten mit einer eindeutigen und einzigartigen Kennung versehen sein. Die Daten sollen über eine Suchfunktion auffindbar sein. Zudem sollen die Daten mit aussagekräftigen Metadaten angereichert sein. Die Metadaten sollen die Kennung der zu beschreibenden Daten beinhalten.

83) Vgl. Universität Duisburg-Essen (2019).

84) Vgl. Eprints (2021).

85) Vgl. Lyrasis (2021a).

86) Vgl. Lyrasis (2021b).

87) Vgl. iRODS Consortiums (2021).

88) Vgl. ASCHENBRENNER/NEUROTH (2011), S. 104.

89) Vgl. ASCHENBRENNER/NEUROTH (2011), S. 105.

90) Vgl. BÜHLER/SCHLAICH/SINNER (2019), S. 81.

91) Vgl. BÜHLER/SCHLAICH/SINNER (2019), S. 81.

92) Vgl. KREUTZER/LAHMANN (2019), S. 74 ff. Auf eine juristische Analyse konkreter Sonderbestimmungen zu Forschungszwecken wird an dieser Stelle verzichtet. Für eine detaillierte Analyse vgl. KREUTZER/LAHMANN (2019), S. 77 ff.

93) Vgl. WILKINSON/DUMONTIER/AALBERSBERG et al. (2016). Neben den FAIR-Prinzipien findet sich in der einschlägigen Fachliteratur eine Vielzahl weiterer Empfehlungen für den Umgang mit Forschungsdaten; vgl. dazu z. B. GOODMAN/PEPE/BLOCKER et al. (2014), S. 1 ff.; WHITE/BALDRIDGE/BRYM et al. (2013), S. 2 ff.

94) Vgl. für die nachfolgenden Bulletpoints WILKINSON/DUMONTIER/AALBERSBERG et al. (2016).

- **Accessible:** Die Daten sollen für Dritte zugänglich sein. Um diesem Prinzip gerecht zu werden, sollen die (Meta-)Daten über ein standardisiertes Kommunikationsprotokoll anhand der Kennung zugänglich sein. Das Kommunikationsprotokoll soll offen, frei und universell implementierbar sein. Zudem soll das Kommunikationsprotokoll eine Authentifizierung und Autorisierung erlauben. Wenn die Daten nicht mehr verfügbar sein sollten, sollen die zugehörigen Metadaten weiterhin zugänglich sein.
- **Interoperable:** Die Daten sollen sich von verschiedenen IT-Systemen nutzen lassen. Um diesem Prinzip gerecht zu werden, soll eine Daten- oder Wissensrepräsentationssprache genutzt werden, die formal, allgemein zugänglich und weit verbreitet ist. Zudem soll das verwendete Vokabular den FAIR-Prinzipien folgen. Die Metadaten sollten darüber hinaus qualifizierte Referenzen zu anderen (Meta-)Daten beinhalten.
- **Reusable:** Die Daten sollten sich wiederverwenden lassen. Um diesem Prinzip gerecht zu werden, sollen die Daten anhand angemessener und relevanter Attribute mittels Metadaten umfangreich beschrieben sein. Die Metadaten sollen mit einer eindeutigen und allgemein zugänglichen Nutzungslizenz veröffentlicht werden. Die Daten sollen außerdem hinsichtlich ihrer Herkunft detailliert beschrieben sein und relevante Standards derjenigen Domäne einhalten, der sie zuzuordnen sind.

Literaturverzeichnis

Vorbemerkungen:

- Alle Quellen werden im Literaturverzeichnis wie folgt aufgeführt: In der ersten Zeile wird der *Referenztitel* der Quelle angegeben. Er entspricht der Form, die im Haupttext Verwendung findet, wenn auf die Quelle hingewiesen wird.
- Bei der Vergabe der Referenztitel wird bei *einem* Autor dessen Nachname, gefolgt von dem Erscheinungsjahr der Quelle in Klammern, verwendet. Existieren *zwei* oder *drei* Autoren, werden diese getrennt von einem Schrägstrich („/“) aufgeführt. Bei mindestens *vier* Autoren werden nur die ersten drei Autoren mit dem Zusatz „et al.“ aufgeführt.
- Zu *Internetquellen* wird die dafür verantwortliche Instanz aufgeführt. Dies können sowohl natürliche als auch juristische Personen sein. Für Internetquellen werden die zum Zugriffsdatum gültige Internetadresse (URL) und das Zugriffsdatum angegeben.

ASCHENBRENNER/NEUROTH (2011)

Aschenbrenner, A.; Neuroth, H.: Forschungsdaten-Repositoryen. In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef 2011, S. 101-114.

BEIERLE/KERN-ISBERNER (2019)

Beierle, C.; Kern-Isberner, G.: Methoden wissensbasierter Systeme – Grundlagen, Algorithmen, Anwendungen. 6. Aufl., Wiesbaden 2019.

BERGENRODT/KOWALSKI (2015)

Bergenrodt, D.; Kowalski, M.: Konzipierung, Implementierung und kritische Evaluierung einer Projektwissensbank auf Basis von semantischen Methoden der Künstlichen Intelligenz. Projektberichte des Verbundprojekts OrGoLo Nr. 31, Institut für Produktion und Industrielles Informationsmanagement, Universität Duisburg-Essen, Campus Essen, Essen 2015.

BERGENRODT/KOWALSKI/ZELEWSKI (2015)

Bergenrodt, D.; Kowalski, M.; Zelewski, S.: Prototypische Implementierung des ontologiegestützten CBR-Tools jCORA. In: Zelewski, S.; Akca, N.; Kowalski, S. (Hrsg.): Organisatorische Innovationen mit Good Governance und Semantic Knowledge Management in Logistik-Netzwerken. Berlin 2015, S. 475-553.

BORCHARDT (2006)

Borchardt, A.: Koordinationsinstrumente in virtuellen Unternehmen – Eine empirische Analyse anhand lose gekoppelter Systeme. Dissertation Universität zu Kiel. Wiesbaden 2006.

BUSBY (1999):

Busby, J. S.: An Assessment of Post-Project Reviews. In: Project Management Journal, Vol. 30 (1999), Issue 3, S. 23-29.

BÜHLER/SCHLAICH/SINNER (2019)

Bühler, P.; Schlaich, P.; Sinner, D.: Datenmanagement – Daten – Datenbanken – Datensicherheit. Berlin 2019.

CASPERS/MOEBUS/LUX et al. (2014)

Caspers, S.; Moebus, S.; Lux, S.; Pundt, N.; Schütz, H.; Mühleisen, T. W.; Gras, V.; Eickhoff, S. B.; Romanzetti, S.; Stöcker, T.; Stirnberg, R.; Kirlangic, M. E.; Minnerop, M.; Pieperhoff, P.; Mödder, U.; Das, S.; Evans, A. C.; Jöckel, K.-H.; Erbel, R.; Cichon, S.; Nöthen, M. M.; Sturma, D.; Bauer, A.; Shah, N. J.; Zilles, K.; Amunts, K.: Studying variability in human brain aging in a populations-based German cohort – rationale and design of 1000BRAINS. In: *Frontiers in Aging Neuroscience*, Vol. 6 (2014), Article 149, S. 809-815.

CHARNES/COOPER (1961)

Charnes, A.; Cooper, W. W.: *Management Models and Industrial Applications of Linear Programming*, Volume 1. New York – London – Sydney 1961.

DANIEL/KUNZ/WAHNIG (2009)

Daniel, K.; Kunz, C.; Wahnig, R.: Wissenssicherung im Projektcontrolling durch Post-Project-Reviews. In: *Controlling – Zeitschrift für erfolgsorientierte Unternehmenssteuerung*, 21. Jg. (2009), Heft 3, S. 183-190.

DINKELBACH (1969)

Dinkelbach, W.: Entscheidungen bei mehrfacher Zielsetzung und die Problematik der Zielgewichtung. In: Busse von Colbe, W.; Meyer-Dohm, P. (Hrsg.): *Unternehmerische Planung und Entscheidung*. Bielefeld 1969, S. 55-70.

ERBEL/EISELE/MOEBUS et al. (2012)

Erbel, R.; Eisele, L.; Moebus, S.; Dragano, N.; Möhlenkamp, S.; Bauer, M.; Kälsch, H.; Jöckel, K.-H.: Die Heinz Nixdorf Recall Studie. In: *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, Vol. 55 (2012), Issue 5-6, S. 809-815.

Eprints (2021):

Eprints: About us. Online-Publikation im Internet unter der URL „<https://www.eprints.org/uk/index.php/about/>“, letzter Zugriff am 17.01.2021.

FREITAG (2011)

Freitag, M.: Projektkommunikation – Eine kritische Betrachtung des Kommunikationsmodells im Projektmanagement. In: Freitag, M.; Müller, C.; Rusch, G.; Spreitzer, T. (Hrsg.): *Projektkommunikation – Strategien für temporäre soziale Systeme*. Wiesbaden 2011, S. 86-174.

FÜRSTENAU/LANGFERMANN/BORN et al. (2005)

Fürstenau, B.; Langfermann, J.; Klauser, F.; Born, V.: Erfahrungswissen sichern und aufbereiten – Zur effizienten Gestaltung von Wissensmanagementprozessen bei der BMW AG im Projekt „Werksaufbau Leipzig“. In: Ferstl, O.; Sinz, E. J.; Eckert, S.; Isselhorst, T. (Hrsg.): *Wirtschaftsinformatik 2005: eEconomy, eGovernment, eSociety*. 7. Internationale Tagung Wirtschaftsinformatik 2005. Heidelberg 2005, S. 1023-1039.

GHOSH (2000)

Ghosh, D. N.: Analytic Hierarchy Process & TOPSIS Method to Evaluate Faculty Performance in Engineering Education. In: *Universal Journal of Applied Computer Science and Technology*, Vol. 1 (2000), Issue 2, S. 63-70.

GOLDBERGER/AMARAL/GLASS et al. (2000)

Goldberger, A.; Amaral, L.; Glass, L.; Hausdorff, P. C.; Ivanov, R.; Mark, J. E.; Mietus, G. B.; Peng, C. K.; Stanley, H. E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. In: *Circulation*, Vol. 101 (2000), Issue 23, S. 215-220.

GOODMAN/PEPE/BLOCKER et al. (2014)

Goodman, A.; Pepe, A.; Blocker, A. W.; Borgman, C. L.; Cranmer, K.; Crosas, M.; Di Stefano, R.; Gil, Y.; Groth, P.; Hedstrom, M.; Hogg, D. W.; Kashyap, V.; Mahabal, A.; Siemiginowska, A.; Slavkovic, A.: Ten Simple Rules for the Care and Feeding of Scientific Data. In: PLOS Computational Biology, Vol. 10 (2014), Issue 4 (ohne Seiten). Online-Publikation im Internet unter der URL „<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003542>“, letzter Zugriff am 23.01.2021.

GÖPFERT/NEHSE (1990)

Göpfert, A.; Nehse, R.: Vektoroptimierung – Theorie, Verfahren, Anwendung. Leipzig 1990.

HEINRICH/KLIER (2018)

Heinrich, B.; Klier, M.: Datenqualitätsmetriken für ein ökonomisch orientiertes Qualitätsmanagement. In: Hildebrand, K.; Gebauer, M.; Hinrichs, H.; Mielke, M. (Hrsg.): Daten- und Informationsqualität – Auf dem Weg zur Information Excellence. 4. Aufl., Wiesbaden 2018, S. 43-65.

HWANG/YOON (1981):

Hwang, C.-L.; Yoon, K.: Multiple Attribute Decision Making – Methods and Applications. A State-of-the-Art Survey. Berlin – Heidelberg 1981.

JUGDEV (2012)

Jugdev, K.: Learning from Lessons Learned: Project Management Research Program. In: American Journal of Economics and Business Administration, Vol. 4 (2012), Issue 1, S. 13-22.

KLUMP (2011)

Klump, J.: Langzeiterhaltung digitaler Forschungsdaten. In: Büttner, S.; Hobohm, H.-C.; Müller, L. (Hrsg.): Handbuch Forschungsdatenmanagement. Bad Honnef 2011, S. 115-122.

KREUTZER/LAHMANN (2019)

Kreutzer, T.; Lahmann, H.: Rechtsfragen bei Open Science: ein Leitfaden. Hamburg 2019.

Lyrasis (2021a):

Lyrasis: Dspace. Online-Publikation im Internet unter der URL „<https://duraspace.org/dspace/>“, letzter Zugriff am 17.01.2021.

Lyrasis (2021b):

Lyrasis: Fedora. Online-Publikation im Internet unter der URL „<https://duraspace.org/fedora/>“, letzter Zugriff am 17.01.2021.

MOODY (1999):

Moody, G.: MIT-BIH Normal Sinus Rhythm Database. Online-Publikation im Internet unter der URL „<https://physionet.org/content/nsrdb/1.0.0/>“, letzter Zugriff am 21.01.2021.

MOODY/MARK (1983):

Moody, G. B.; Mark, R. G.: An new method for detecting atrial fibrillation using R-R intervals. In: Computers in Cardiology, Vol. 20 (1983), No. 3, S. 45-50.

MOODY/MARK (2001):

Moody, G. B.; Mark, R. G.: The Impact of the MIT-BIH Arrhythmia Database. In: IEEE Engineering in Medicine and Biology, Vol. 20 (2001), Issue 3, S. 45-50.

MOODY/MARK/GOLDBERGER (1988):

Moody, G. B.; Mark, R. G.; Goldberger, A. L.: Evaluation of the „TRIM“ ECG data compressor. In: Computers in Cardiology, Vol. 15 (1988), S. 167-170.

MOODY/MULDROW/MARK (1984):

Moody, G. B.; Muldrow, W. E.; Mark, R. G.: A noise stress test for arrhythmia detectors. In: *Computers in Cardiology*, Vol. 11 (1984), S. 381-384.

NONAKA/TAKEUCHI (2012):

Nonaka, I.; Takeuchi, H.: *Die Organisation des Wissens – Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*. 2. Aufl., Frankfurt am Main 2012.

PETERS/SCHÜTTE/ZELEWSKI (2006):

Peters, M.; Schütte, R.; Zelewski, S.: *Erweiterte Wirtschaftlichkeitsanalyse mithilfe des Analytic Hierarchy Process (AHP) unter Berücksichtigung des Wissensmanagements zur Beurteilung von Filialen eines Handelsunternehmens*. Institut für Produktion und Industrielles Informationsmanagement, Arbeitsbericht Nr. 30. Universität Duisburg-Essen (Campus Essen), Essen 2006.

PIRO/GEBAUER (2018)

Piro, A.; Gebauer, M.: *Definition von Datenarten zur konsistenten Kommunikation im Unternehmen*. In: Hildebrand, K.; Gebauer, M.; Hinrichs, H.; Mielke, M. (Hrsg.): *Daten- und Informationsqualität – Auf dem Weg zur Information Excellence*. 4. Aufl., Wiesbaden 2018, S. 141-154.

POLANYI (2009):

Polanyi, M.: *The Tacit Dimension*. Chicago - London 2009.

re3data.org (2021):

re3data.org: About. Online-Publikation im Internet unter der URL „<https://www.re3data.org/about>“, letzter Zugriff am 21.01.2021.

RIEDL (2006):

Riedl, R.: *Analytischer Hierarchieprozess vs. Nutzwertanalyse: Eine vergleichende Gegenüberstellung zweier multiattributiver Auswahlverfahren am Beispiel Application Service Providing*. In: Fink, K.; Ploder, C. (Hrsg.): *Wirtschaftsinformatik als Schlüssel zum Unternehmenserfolg*. Wiesbaden 2006, S. 99-127.

ROHWEDER/KASTEN/MALZAHN et al. (2018)

Rohweder, J. P.; Kasten, G.; Malzahn, D.; Piro, A.; Schmid, J.: *Informationsqualität – Definitionen, Dimensionen und Begriffe*. In: Hildebrand, K.; Gebauer, M.; Hinrichs, H.; Mielke, M. (Hrsg.): *Daten- und Informationsqualität – Auf dem Weg zur Information Excellence*. 4. Aufl., Wiesbaden 2018, S. 23-43.

SAATY (1986):

Saaty, T. L.: *Axiomatic Foundation of the Analytic Hierarchy Process*. In: *Management Science*, Vol. 32 (1986), No. 7, S. 841-855.

SAATY (1990):

Saaty, T. L.: *How to make a decision: The Analytic Hierarchy Process*. In: *European Journal of Operational Research*, Vol. 48 (1990), Issue 1, S. 9-26.

SAATY (1994):

Saaty, T. L.: *How to Make a Decision: The Analytic Hierarchy Process*. In: *Interfaces*, Vol. 24 (1994), No. 6, S. 19-43.

SAATY (2000):

Saaty, T. L.: *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process*, Vol. 6. 2. Aufl., Pittsburgh 2000.

SAATY (2008):

Saaty, T. L.: *Decision Making for Leaders – The Analytic Hierarchy Process for Decisions in a Complex World*. Pittsburgh 2008.

SAATY (2005):

Saaty, T. L.: Making and validating complex decisions with the AHP/ANP. In: Journal of Systems Science and Systems Engineering, Vol. 14 (2005), Issue 1, S. 1-36.

SAATY/VARGAS (1998):

Saaty, T. L.; Vargas, L. G.: Diagnosis with Dependent Symptoms: Bayes Theorem and the Analytic Hierarchy Process. In: Operations Research, Vol. 46 (1998), No. 4, S. 491-502.

SCHAGEN/ZELEWSKI/HEEB (2020):

Schagen, J. P.; Zelewski, S.; Heeb, T.: Erhebung und Analyse der Anforderungen an ein KI-Tool aus der Perspektive der betrieblichen Praxis – mit Fokus auf der Wiederverwendung von Erfahrungswissen im Bereich des betrieblichen Projektmanagements. Institut für Produktion und Industrielles Informationsmanagement, Arbeitsbericht Nr. 47, zugleich KI-LiveS-Projektbericht Nr. 1. Universität Duisburg-Essen (Campus Essen), Essen 2020.

SCHINDLER/EPPLER (2003):

Schindler, M.; Eppler, M. J.: Harvesting Project Knowledge: A Review of Project Learning Methods and Success Factors. In: International Journal of Project Management, Vol. 21 (2003), No. 3, S. 219-228.

SCHNEEWEIB (1990):

Schneeweiß, C.: Kostenwirksamkeitsanalyse, Nutzwertanalyse und Multi-Attributive Nutzen-theorie. In: Wirtschaftswissenschaftliches Studium, 19. Jg. (1990), Heft 1, S. 13-18.

STORCH/LAUE/GRUHN (2013):

Storch, A.; Laue, R.; Gruhn, V.: Measuring and Visualising the Quality of Models. In: IEEE: 2013 IEEE 1st International Workshop on Communicating Business Process and Software Models Quality, Understandability, and Maintainability (CPSM), Eindhoven. Piscataway 2013, S. 1-8.

SUEUR (2018):

Sueur, J.: Sound Analysis and Synthesis with R. Cham 2018.

Universität Duisburg-Essen (2019):

Universität Duisburg-Essen: Speicher-Matrix. Online-Publikation im Internet unter der URL „https://www.uni-due.de/rds/speicher_matrix.php“, letzte Änderung am 19.12.2019, letzter Zugriff am 21.01.2021.

Universität Duisburg-Essen (2020):

Universität Duisburg-Essen: Was ist Open Access? Online-Publikation im Internet unter der URL „https://www.uni-due.de/ub/publikationsdienste/oa_definition.php“, letzte Änderung am 01.07.2020, letzter Zugriff am 17.01.2021.

WANG/LEE (2007)

Wang, Y.-J.; Lee, H.-S.: Generalizing TOPSIS for fuzzy multiple-criteria group decision making. In: Computers and Mathematics with Applications, Vol. 53 (2007), Issue 11, S. 1762-1772.

WANG/LU/WANG et al. (2017):

Wang, L.; Lu, H.; Wang, Y.; Feng, M.; Wang, D.; Yin, B.; Ruan, X.: Learning to Detect Salient Objects with Image-level Supervision. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu. Piscataway 2017, S. 3796-3805.

WENGER (2000):

Wenger, E.: Communities of Practice and Social Learning Systems. In: Organization, Vol. 7 (2000), Issue 2, S. 225-246.

WHITE/BALDRIDGE/BRYM et al. (2013):

White, E. P.; Baldrige, E.; Brym, Z. T.; Locey, K. J.; McGlenn, D. J.; Supp, S. R.: Nine simple ways to make it easier to (re)use your data. In: *Ideas in Ecology and Evolution*, Vol. 6 (2013), No. 2, S. 1-10.

WILKINSON/DUMONTIER/AALBERSBERG et al. (2016):

Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; Bonino da Silva Santos, L.; Bourne, P. E.; Bouwman, J.; Brookes, A.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Marone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S.-A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data*, Vol. 3 (2016), ohne Seiten. Online-Publikation im Internet unter der URL „<https://www.nature.com/articles/sdata201618>“, letzter Zugriff am 23.01.2021.

WOLF (2018)

Wolf, J.: Organisatorische Maßnahmen für gute Datenqualität. In: Hildebrand, K.; Gebauer, M.; Hinrichs, H.; Mielke, M. (Hrsg.): *Daten- und Informationsqualität – Auf dem Weg zur Information Excellence*. 4. Aufl., Wiesbaden 2018, S. 235-252.

ZANGEMEISTER (1976)

Zangemeister, C.: *Nutzwertanalyse in der Systemtechnik – Eine Methodik zur multidimensionalen Bewertung und Auswahl von Projektalternativen*. 4. Aufl., München 1976.

ZELEWSKI/PETERS (2005)

Zelewski, S./ Peters, M. L.: Goal Programming zur Lösung von Zuordnungsproblemen. In: *Das Wirtschaftsstudium*, 34. Jg. (2005), Heft 10, S. 1208-1218.

Anhang A: Leitfaden Experteninterview

Leitfaden zur Datenerfassung im KI-LiveS-Projekt

Die nachfolgenden Fragen beziehen sich nur auf Daten (Dateien), die Sie bereit wären, im Rahmen des KI-LiveS-Projekts, das vom BMBF gefördert wird, als „Datenquellen“ für ein „Data Repository“ zur Verfügung zu stellen. Das BMBF wünscht sich ein solches „Data Repository“, um in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für weitere Entwicklungsarbeiten im KI-Umfeld zu unterstützen. Selbstverständlich sind uns die vitalen Interessen von Unternehmen bewusst, ihre wettbewerbsrelevanten Daten zu schützen. Daher bitten wir Sie, uns von vornherein wissen zu lassen, in welchem Ausmaß Sie zu einer Kooperation mit dem BMBF-KI-LiveS-Projekt hinsichtlich der Zurverfügungstellung von Daten bereit sind. Die von Ihnen bereitgestellten Daten können Ihren Vorgaben entsprechend anonymisiert werden.

Bereitschaft zur Datenbereitstellung: Wählen Sie ein Element aus.

Die Beantwortung der nachfolgenden Fragen erübrigt sich, falls Sie zuvor die Option „keine Bereitschaft zur Verfügungstellung von Daten“ gewählt haben. Wir würden dies vollkommen verstehen und entsprechend respektieren.

Allgemeine Angaben zu den Daten

Datum der Datenerfassung:

Herkunft der Daten:

Wählen Sie ein Element aus.

bitte ergänzen, falls „Sonstiges“:

Schlüsselwörter:

Technische Angaben zu den Daten

Sind die Daten in eine *bekannte Kategorie* einzuordnen? (Mehrfachnennungen möglich):

- Sensordaten: Wählen Sie ein Element aus.

Sonstiges:

bitte ergänzen:

- Tabellendaten:

- Textdaten:

Wählen Sie ein Element aus.

Sonstiges:

bitte ergänzen:

- Sonstiges:

bitte ergänzen:

Liegen die Daten in einem *standardisierten Format* vor? (Mehrfachnennungen möglich)

- Bilddaten: Wählen Sie ein Element aus.

Sonstiges:

- Videodaten:

Wählen Sie ein Element aus.

Sonstiges:

- Audiodaten:

Wählen Sie ein Element aus.

Sonstiges:

- Tabellendaten:

Wählen Sie ein Element aus.

Sonstiges:

- Textdaten:

Wählen Sie ein Element aus.

Sonstiges:

- weitere Datenformate:

Strukturierung der Daten

Sind die Daten strukturiert?

Wählen Sie ein Element aus.

Sind die Daten für überwachte Lernverfahren geeignet?

Wählen Sie ein Element aus.

Ist die Problemstellung klar formuliert?

Wählen Sie ein Element aus.

Quantität der Daten

Wie groß ist der *durchschnittliche Umfang* Ihrer Dateien (*grob geschätzt* in Mbyte oder GByte):

- Bilddaten: _____ MByte GByte
- Videodateien: _____ MByte GByte
- Audiodateien: _____ MByte GByte
- Tabellendaten: _____ MByte GByte
- Textdateien: _____ MByte GByte
- Sonstiges: _____ MByte GByte

Wie *viele Dateien* verwenden Sie in etwa in Ihrem Unternehmen für die Unterstützung Ihrer Geschäftsprozesse *pro Jahr* (Anzahl *grob geschätzt*):

- Bilddateien: Wählen Sie ein Element aus.
- Videodateien: Wählen Sie ein Element aus.
- Audiodateien: Wählen Sie ein Element aus.
- Tabellendateien: Wählen Sie ein Element aus.
- Textdateien: Wählen Sie ein Element aus.
- Sonstiges: Wählen Sie ein Element aus.

Qualität der Daten

Wie hoch ist der Anteil *annotierter* Daten? Wählen Sie ein Element aus.

Wie *homogen* sind die vorhandenen Daten? Wählen Sie ein Element aus.

Sind die Daten bereits in irgendeiner Form *vorverarbeitet* worden?

- Sensordaten: die Sensordaten kommen direkt vom Gerät
 die Sensordaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Tabellendaten: die Tabellendaten werden direkt aus Tabellendateien entnommen
 die Tabellendaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Textdaten: die Textdaten werden direkt aus Textdateien entnommen
 die Textdaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____

In welche Kategorie fallen die intendierten Anwendungen der von Ihnen vorgehaltenen Daten?
(Mehrfachnennungen möglich)

- Klassifikation: Einteilung in eine von mehreren vorgegebenen Klassen
 - Segmentierung (insbesondere Bild- und Videodaten):
Einteilung in zusammenhängende Bereiche
 - Detektion/Lokalisation (insbesondere Bild- und Videodaten):
Bestimmung relevanter Positionen
 - Wiederverwendung von projektbezogenem Erfahrungswissen
 - Sonstiges:
-

Anhang B: Ausgefüllte Leitfäden zur Datenerfassung im KI-LiveS-Projekt

Ausgefüllter Leitfaden zur Datenerfassung im KI-LiveS-Projekt

Die nachfolgenden Fragen beziehen sich nur auf Daten (Dateien), die Sie bereit wären, im Rahmen des KI-LiveS-Projekts, das vom BMBF gefördert wird, als „Datenquellen“ für ein „Data Repository“ zur Verfügung zu stellen. Das BMBF wünscht sich ein solches „Data Repository“, um in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für weitere Entwicklungsarbeiten im KI-Umfeld zu unterstützen. Selbstverständlich sind uns die vitalen Interessen von Unternehmen bewusst, ihre wettbewerbsrelevanten Daten zu schützen. Daher bitten wir Sie, uns von vornherein wissen zu lassen, in welchem Ausmaß Sie zu einer Kooperation mit dem BMBF-KI-LiveS-Projekt hinsichtlich der Zurverfügungstellung von Daten bereit sind. Die von Ihnen bereitgestellten Daten können Ihren Vorgaben entsprechend anonymisiert werden.

Bereitschaft zur Datenbereitstellung:

Bereitschaft zur Verfügungstellung von Daten für BMBF-geförderte Projekte

Die Beantwortung der nachfolgenden Fragen erübrigt sich, falls Sie zuvor die Option „keine Bereitschaft zur Verfügungstellung von Daten“ gewählt haben. Wir würden dies vollkommen verstehen und entsprechend respektieren.

Allgemeine Angaben zu den Daten

Datum der Datenerfassung: Mai 2019 – Mai 2020
 Herkunft der Daten: Sonstiges
 bitte ergänzen, falls „Sonstiges“: Wissenschaftliche Daten
 Schlüsselwörter: Urban Sounds, Audioaufnahmen

Technische Angaben zu den Daten

Sind die Daten in eine *bekannte Kategorie* einzuordnen? (Mehrfachnennungen möglich):

- Sensordaten: Audiodaten
 Sonstiges:
- bitte ergänzen: _____
- Tabellendaten:
- Textdaten: Wählen Sie ein Element aus.
 Sonstiges:
- bitte ergänzen: _____
- Sonstiges:
- bitte ergänzen: _____

Liegen die Daten in einem *standardisierten Format* vor? (Mehrfachnennungen möglich)

- Bilddaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Videodaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Audiodaten: WAV
 Sonstiges: _____
- Tabellendaten: Excel
 Sonstiges: _____

- Textdaten: Wählen Sie ein Element aus.
Sonstiges: _____
- weitere Datenformate: _____

Strukturierung der Daten

- Sind die Daten strukturiert? ja
Sind die Daten für überwachte Lernverfahren geeignet? ja
Ist die Problemstellung klar formuliert? nein

Quantität der Daten

Wie groß ist der *durchschnittliche Umfang* Ihrer Dateien (*grob geschätzt* in Mbyte oder GByte):

- Bilddaten: _____ MByte GByte
- Videodateien: _____ MByte GByte
- Audiodaten: 10000 _____ MByte GByte
- Tabellendaten: 10 _____ MByte GByte
- Textdaten: _____ MByte GByte
- Sonstiges: _____ MByte GByte

Wie *viele Dateien* verwenden Sie in etwa in Ihrem Unternehmen für die Unterstützung Ihrer Geschäftsprozesse *pro Jahr* (Anzahl grob geschätzt):

- Bilddateien: Wählen Sie ein Element aus.
- Videodateien: Wählen Sie ein Element aus.
- Audiodateien: 10.000 und mehr
- Tabellendateien: weniger als 100
- Textdateien: Wählen Sie ein Element aus.
- Sonstiges: Wählen Sie ein Element aus.

Qualität der Daten

Wie hoch ist der Anteil *annotierter* Daten? keine Aussage

Wie *homogen* sind die vorhandenen Daten? hoch

Sind die Daten bereits in irgendeiner Form *vorverarbeitet* worden?

- Sensordaten: die Sensordaten kommen direkt vom Gerät
 die Sensordaten wurden nachbearbeitet
falls ja, in welcher Weise: _____
- Tabellendaten: die Tabellendaten werden direkt aus Tabellendateien entnommen
 die Tabellendaten wurden nachbearbeitet
falls ja, in welcher Weise: Ergänzung um berechnete Variablen
(Sound-Indizes sowie bebaute Umwelt)
- Textdaten: die Textdaten werden direkt aus Textdateien entnommen
 die Textdaten wurden nachbearbeitet
falls ja, in welcher Weise: _____

In welche Kategorie fallen die intendierten Anwendungen der von Ihnen vorgehaltenen Daten?
(Mehrfachnennungen möglich)

- Klassifikation: Einteilung in eine von mehreren vorgegebenen Klassen
 - Segmentierung (insbesondere Bild- und Videodaten):
Einteilung in zusammenhängende Bereiche
 - Detektion/Lokalisation (insbesondere Bild- und Videodaten):
Bestimmung relevanter Positionen
 - Wiederverwendung von projektbezogenem Erfahrungswissen
 - Sonstiges:
-

Ausgefüllter Leitfaden zur Datenerfassung im KI-LiveS-Projekt

Die nachfolgenden Fragen beziehen sich nur auf Daten (Dateien), die Sie bereit wären, im Rahmen des KI-LiveS-Projekts, das vom BMBF gefördert wird, als „Datenquellen“ für ein „Data Repository“ zur Verfügung zu stellen. Das BMBF wünscht sich ein solches „Data Repository“, um in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für weitere Entwicklungsarbeiten im KI-Umfeld zu unterstützen. Selbstverständlich sind uns die vitalen Interessen von Unternehmen bewusst, ihre wettbewerbsrelevanten Daten zu schützen. Daher bitten wir Sie, uns von vornherein wissen zu lassen, in welchem Ausmaß Sie zu einer Kooperation mit dem BMBF-KI-LiveS-Projekt hinsichtlich der Zurverfügungstellung von Daten bereit sind. Die von Ihnen bereitgestellten Daten können Ihren Vorgaben entsprechend anonymisiert werden.

Bereitschaft zur Datenbereitstellung: keine Bereitschaft zur Verfügungstellung von Daten

Die Beantwortung der nachfolgenden Fragen erübrigt sich, falls Sie zuvor die Option „keine Bereitschaft zur Verfügungstellung von Daten“ gewählt haben. Wir würden dies vollkommen verstehen und entsprechend respektieren.

Allgemeine Angaben zu den Daten

Datum der Datenerfassung: _____
 Herkunft der Daten: Wählen Sie ein Element aus.
 bitte ergänzen, falls „Sonstiges“: _____
 Schlüsselwörter: _____

Technische Angaben zu den Daten

Sind die Daten in eine *bekannte Kategorie* einzuordnen? (Mehrfachnennungen möglich):

- Sensordaten: Wählen Sie ein Element aus.
 Sonstiges:
 bitte ergänzen: _____
- Tabellendaten:
- Textdaten: Wählen Sie ein Element aus.
 Sonstiges:
 bitte ergänzen: _____
- Sonstiges:
 bitte ergänzen: _____

Liegen die Daten in einem *standardisierten Format* vor? (Mehrfachnennungen möglich)

- Bilddaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Videodaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Audiodaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Tabellendaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Textdaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- weitere Datenformate: _____

Strukturierung der Daten

Sind die Daten strukturiert?

Wählen Sie ein Element aus.

Sind die Daten für überwachte Lernverfahren geeignet?

Wählen Sie ein Element aus.

Ist die Problemstellung klar formuliert?

Wählen Sie ein Element aus.

Quantität der Daten

Wie groß ist der *durchschnittliche Umfang* Ihrer Dateien (*grob geschätzt* in Mbyte oder GByte):

- Bilddaten: _____ MByte GByte
- Videodaten: _____ MByte GByte
- Audiodaten: _____ MByte GByte
- Tabellendaten: _____ MByte GByte
- Textdaten: _____ MByte GByte
- Sonstiges: _____ MByte GByte

Wie *viele Dateien* verwenden Sie in etwa in Ihrem Unternehmen für die Unterstützung Ihrer Geschäftsprozesse *pro Jahr* (Anzahl *grob geschätzt*):

- Bilddateien: Wählen Sie ein Element aus.
- Videodateien: Wählen Sie ein Element aus.
- Audiodateien: Wählen Sie ein Element aus.
- Tabellendateien: Wählen Sie ein Element aus.
- Textdateien: Wählen Sie ein Element aus.
- Sonstiges: Wählen Sie ein Element aus.

Qualität der Daten

Wie hoch ist der Anteil *annotierter* Daten? Wählen Sie ein Element aus.

Wie *homogen* sind die vorhandenen Daten? Wählen Sie ein Element aus.

Sind die Daten bereits in irgendeiner Form *vorverarbeitet* worden?

- Sensordaten: die Sensordaten kommen direkt vom Gerät
 die Sensordaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Tabellendaten: die Tabellendaten werden direkt aus Tabellendateien entnommen
 die Tabellendaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Textdaten: die Textdaten werden direkt aus Textdateien entnommen
 die Textdaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____

In welche Kategorie fallen die intendierten Anwendungen der von Ihnen vorgehaltenen Daten?
(Mehrfachnennungen möglich)

- Klassifikation: Einteilung in eine von mehreren vorgegebenen Klassen
 - Segmentierung (insbesondere Bild- und Videodaten):
Einteilung in zusammenhängende Bereiche
 - Detektion/Lokalisation (insbesondere Bild- und Videodaten):
Bestimmung relevanter Positionen
 - Wiederverwendung von projektbezogenem Erfahrungswissen
 - Sonstiges:
-

Ausgefüllter Leitfaden zur Datenerfassung im KI-LiveS-Projekt

Die nachfolgenden Fragen beziehen sich nur auf Daten (Dateien), die Sie bereit wären, im Rahmen des KI-LiveS-Projekts, das vom BMBF gefördert wird, als „Datenquellen“ für ein „Data Repository“ zur Verfügung zu stellen. Das BMBF wünscht sich ein solches „Data Repository“, um in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für weitere Entwicklungsarbeiten im KI-Umfeld zu unterstützen. Selbstverständlich sind uns die vitalen Interessen von Unternehmen bewusst, ihre wettbewerbsrelevanten Daten zu schützen. Daher bitten wir Sie, uns von vornherein wissen zu lassen, in welchem Ausmaß Sie zu einer Kooperation mit dem BMBF-KI-LiveS-Projekt hinsichtlich der Zurverfügungstellung von Daten bereit sind. Die von Ihnen bereitgestellten Daten können Ihren Vorgaben entsprechend anonymisiert werden.

Bereitschaft zur Datenbereitstellung: keine Bereitschaft zur Verfügungstellung von Daten

Die Beantwortung der nachfolgenden Fragen erübrigt sich, falls Sie zuvor die Option „keine Bereitschaft zur Verfügungstellung von Daten“ gewählt haben. Wir würden dies vollkommen verstehen und entsprechend respektieren.

Allgemeine Angaben zu den Daten

Datum der Datenerfassung: _____
 Herkunft der Daten: Wählen Sie ein Element aus.
 bitte ergänzen, falls „Sonstiges“: _____
 Schlüsselwörter: _____

Technische Angaben zu den Daten

Sind die Daten in eine *bekannte Kategorie* einzuordnen? (Mehrfachnennungen möglich):

- Sensordaten: Wählen Sie ein Element aus.
 Sonstiges:
 bitte ergänzen: _____
- Tabellendaten:
- Textdaten: Wählen Sie ein Element aus.
 Sonstiges:
 bitte ergänzen: _____
- Sonstiges:
 bitte ergänzen: _____

Liegen die Daten in einem *standardisierten Format* vor? (Mehrfachnennungen möglich)

- Bilddaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Videodaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Audiodaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Tabellendaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- Textdaten: Wählen Sie ein Element aus.
 Sonstiges: _____
- weitere Datenformate: _____

Strukturierung der Daten

Sind die Daten strukturiert?

Wählen Sie ein Element aus.

Sind die Daten für überwachte Lernverfahren geeignet?

Wählen Sie ein Element aus.

Ist die Problemstellung klar formuliert?

Wählen Sie ein Element aus.

Quantität der Daten

Wie groß ist der *durchschnittliche Umfang* Ihrer Dateien (*grob geschätzt* in Mbyte oder GByte):

- Bilddaten: _____ MByte GByte
- Videodaten: _____ MByte GByte
- Audiodaten: _____ MByte GByte
- Tabellendaten: _____ MByte GByte
- Textdaten: _____ MByte GByte
- Sonstiges: _____ MByte GByte

Wie *viele Dateien* verwenden Sie in etwa in Ihrem Unternehmen für die Unterstützung Ihrer Geschäftsprozesse *pro Jahr* (Anzahl *grob geschätzt*):

- Bilddateien: Wählen Sie ein Element aus.
- Videodateien: Wählen Sie ein Element aus.
- Audiodateien: Wählen Sie ein Element aus.
- Tabellendateien: Wählen Sie ein Element aus.
- Textdateien: Wählen Sie ein Element aus.
- Sonstiges: Wählen Sie ein Element aus.

Qualität der Daten

Wie hoch ist der Anteil *annotierter* Daten? Wählen Sie ein Element aus.

Wie *homogen* sind die vorhandenen Daten? Wählen Sie ein Element aus.

Sind die Daten bereits in irgendeiner Form *vorverarbeitet* worden?

- Sensordaten: die Sensordaten kommen direkt vom Gerät
 die Sensordaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Tabellendaten: die Tabellendaten werden direkt aus Tabellendateien entnommen
 die Tabellendaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Textdaten: die Textdaten werden direkt aus Textdateien entnommen
 die Textdaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____

In welche Kategorie fallen die intendierten Anwendungen der von Ihnen vorgehaltenen Daten?
(Mehrfachnennungen möglich)

- Klassifikation: Einteilung in eine von mehreren vorgegebenen Klassen
 - Segmentierung (insbesondere Bild- und Videodaten):
Einteilung in zusammenhängende Bereiche
 - Detektion/Lokalisation (insbesondere Bild- und Videodaten):
Bestimmung relevanter Positionen
 - Wiederverwendung von projektbezogenem Erfahrungswissen
 - Sonstiges:
-

Ausgefüllter Leitfaden zur Datenerfassung im KI-LiveS-Projekt

Die nachfolgenden Fragen beziehen sich nur auf Daten (Dateien), die Sie bereit wären, im Rahmen des KI-LiveS-Projekts, das vom BMBF gefördert wird, als „Datenquellen“ für ein „Data Repository“ zur Verfügung zu stellen. Das BMBF wünscht sich ein solches „Data Repository“, um in Zeiten von „Big Data“ ein möglichst umfassendes Datenfundament für weitere Entwicklungsarbeiten im KI-Umfeld zu unterstützen. Selbstverständlich sind uns die vitalen Interessen von Unternehmen bewusst, ihre wettbewerbsrelevanten Daten zu schützen. Daher bitten wir Sie, uns von vornherein wissen zu lassen, in welchem Ausmaß Sie zu einer Kooperation mit dem BMBF-KI-LiveS-Projekt hinsichtlich der Zurverfügungstellung von Daten bereit sind. Die von Ihnen bereitgestellten Daten können Ihren Vorgaben entsprechend anonymisiert werden.

Bereitschaft zur Datenbereitstellung:

Die Beantwortung der nachfolgenden Fragen erübrigt sich, falls Sie zuvor die Option „keine Bereitschaft zur Verfügungstellung von Daten“ gewählt haben. Wir würden dies vollkommen verstehen und entsprechend respektieren.

Allgemeine Angaben zu den Daten

Datum der Datenerfassung:

Herkunft der Daten:

bitte ergänzen, falls „Sonstiges“:

Schlüsselwörter:

Technische Angaben zu den Daten

Sind die Daten in eine *bekannte Kategorie* einzuordnen? (Mehrfachnennungen möglich):

- Sensordaten:
Sonstiges:
bitte ergänzen:
- Tabellendaten:
- Textdaten:
Sonstiges:
bitte ergänzen:
- Sonstiges:
bitte ergänzen:

Liegen die Daten in einem *standardisierten Format* vor? (Mehrfachnennungen möglich)

- Bilddaten:
Sonstiges:
- Videodaten:
Sonstiges:
- Audiodaten:
Sonstiges:
- Tabellendaten:
Sonstiges:
- Textdaten:
Sonstiges:
- weitere
Datenformate:

Strukturierung der Daten

Sind die Daten strukturiert?

Sind die Daten für überwachte Lernverfahren geeignet?

Ist die Problemstellung klar formuliert?

Quantität der Daten

Wie groß ist der *durchschnittliche Umfang* Ihrer Dateien (*grob geschätzt* in MByte oder GByte):

- Bilddaten: _____ MByte GByte
- Videodateien: _____ MByte GByte
- Audiodaten: _____ MByte GByte
- Tabellendaten: _____ MByte GByte
- Textdaten: _____ MByte GByte
- Sonstiges: _____ MByte GByte

Wie *viele Dateien* verwenden Sie in etwa in Ihrem Unternehmen für die Unterstützung Ihrer Geschäftsprozesse *pro Jahr* (Anzahl *grob geschätzt*):

- Bilddateien:
- Videodateien:
- Audiodateien:
- Tabellendateien:
- Textdateien:
- Sonstiges:

Qualität der Daten

Wie hoch ist der Anteil *annotierter* Daten?

Wie *homogen* sind die vorhandenen Daten?

Sind die Daten bereits in irgendeiner Form *vorverarbeitet* worden?

- Sensordaten: die Sensordaten kommen direkt vom Gerät
 die Sensordaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Tabellendaten: die Tabellendaten werden direkt aus Tabellendateien entnommen
 die Tabellendaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____
- Textdaten: die Textdaten werden direkt aus Textdateien entnommen
 die Textdaten wurden nachbearbeitet
 falls ja, in welcher Weise: _____

In welche Kategorie fallen die intendierten Anwendungen der von Ihnen vorgehaltenen Daten?
 (Mehrfachnennungen möglich)

- Klassifikation: Einteilung in eine von mehreren vorgegebenen Klassen
- Segmentierung (insbesondere Bild- und Videodateien): Einteilung in zusammenhängende Bereiche
- Detektion/Lokalisation (insbesondere Bild- und Videodateien): Bestimmung relevanter Positionen
- Wiederverwendung von projektbezogenem Erfahrungswissen
- Sonstiges: _____

**Institut für Produktion und
Industrielles Informationsmanagement
Universität Duisburg-Essen / Campus Essen**

**Verzeichnis der Arbeitsberichte
(ISSN 1614-0842)**

- Nr. 1: Zelewski, S.: Stickels theoretische Begründung des Produktivitätsparadoxons der Informationstechnik. Universität Essen, Essen 1999.
- Nr. 2: Zelewski, S.: Flexibilitätsorientierte Koordinierung von Produktionsprozessen. Universität Essen, Essen 1999.
- Nr. 3: Zelewski, S.: Ontologien zur Strukturierung von Domänenwissen. Universität Essen, Essen 1999.
- Nr. 4: Siedentopf, J.; Schütte, R.; Zelewski, S.: Wirtschaftsinformatik und Wissenschaftstheorie. Universität Essen, Essen 1999.
- Nr. 5: Fischer, K.; Zelewski, S.: Ontologiebasierte Koordination von Anpassungsplanungen in Produktions- und Logistiknetzwerken mit Multi-Agenten-Systemen. Universität Essen, Essen 1999.
- Nr. 6: Weihermann, A. E.; Wöhlert, K.: Gentechnikakzeptanz und Kommunikationsmaßnahmen in der Lebensmittelindustrie. Universität Essen, Essen 1999.
- Nr. 7: Schütte, R.: Zum Realitätsbezug von Informationsmodellen. Universität Essen, Essen 2000.
- Nr. 8: Zelewski, S.: Erweiterungen eines Losgrößenmodells für betriebliche Entsorgungsprobleme. Universität Essen, Essen 2000.
- Nr. 9: Schütte, R.: Wissen, Zeichen, Information, Daten. Universität Essen, Essen 2000.
- Nr. 10: Hemmert, M.: The Impact of Internationalization and Externalization on the Technology Acquisition Performance of High-Tech Firms. Universität Essen, Essen 2001.
- Nr. 11: Hemmert, M.: Erfolgswirkungen der internationalen Organisation von Technologiegewinnungsaktivitäten. Universität Essen, Essen 2001.
- Nr. 12: Hemmert, M.: Erfolgsfaktoren der Technologiegewinnung von F&E-intensiven Großunternehmen. Universität Essen, Essen 2001.
- Nr. 13: Schütte, R.; Zelewski, S.: Epistemological Problems in Working with Ontologies. Universität Essen, Essen 2001.
- Nr. 14: Peters, M. L.; Zelewski, S.: Analytical Hierarchy Process (AHP). Universität Essen, Essen 2002.
- Nr. 15: Zelewski, S.: Wissensmanagement mit Ontologien. Universität Essen, Essen 2002.
- Nr. 16: Klumpp, M.; Krol, B.; Zug, S.: Management von Kompetenzprofilen im Gesundheitswesen. Universität Essen, Essen 2002.
- Nr. 17: Zelewski, S.: Der „non statement view“ – eine Herausforderung für die (Re-) Konstruktion wirtschaftswissenschaftlicher Theorien. Universität Essen, Essen 2002.
- Nr. 18: Peters, M. L.; Zelewski, S.: A heuristic algorithm to improve the consistency of judgments in the Analytical Hierarchy Process (AHP). Universität Duisburg-Essen (Campus Essen), Essen 2003.

- Nr. 19: Peters, M. L.; Zelewski, S.: Fallstudie zur Lösung eines Standortplanungsproblems mit Hilfe des Analytical Hierarchy Process (AHP). Universität Duisburg-Essen (Campus Essen), Essen 2003.
- Nr. 20: Zelewski, S.: Konventionelle versus strukturalistische Produktionstheorie. Universität Duisburg-Essen (Campus Essen), Essen 2003.
- Nr. 21: Alparslan, A.; Zelewski, S.: Moral Hazard in JIT Production Settings. Universität Duisburg-Essen (Campus Essen), Essen 2004.
- Nr. 22: Dittmann, L.: Ontology-based Skills Management. Universität Duisburg-Essen (Campus Essen), Essen 2004.
- Nr. 23: Peters, M. L.; Zelewski, S.: Ein Modell zur Auswahl von Produktionsaufträgen unter Berücksichtigung von Synergien. Universität Duisburg-Essen (Campus Essen), Essen 2004.
- Nr. 24: Peters, M. L.; Zelewski, S.: Ein Modell zur Zuordnung ähnlicher Kundenbetreuer zu Kunden. Universität Duisburg-Essen (Campus Essen), Essen 2004.
- Nr. 25: Zelewski, S.: Kooperatives Wissensmanagement in Engineering-Netzwerken – (vorläufiger) Abschlussbericht zum Verbundprojekt KOWIEN. Universität Duisburg-Essen (Campus Essen), Essen 2004.
- Nr. 26: Siemens, F.: Vorgehensmodell zur Auswahl einer Variante der Data Envelopment Analysis. Universität Duisburg-Essen (Campus Essen), Essen 2005.
- Nr. 27: Alan, Y.: Integrative Modellierung kooperativer Informationssysteme – Ein Konzept auf der Basis von Ontologien und Petri-Netzen. Dissertation, Universität Duisburg-Essen (Campus Essen), Essen 2005.
- Nr. 28: Akca, N.; Ilas, A.: Produktionsstrategien – Überblick und Systematisierung. Universität Duisburg-Essen (Campus Essen), Essen 2005.
- Nr. 29: Zelewski, S.: Relativer Fortschritt von Theorien – ein strukturalistisches Rahmenkonzept zur Beurteilung der Fortschrittlichkeit wirtschaftswissenschaftlicher Theorien (Langfassung). Universität Duisburg-Essen (Campus Essen), Essen 2005.
- Nr. 30: Peters, M. L.; Schütte, R.; Zelewski, S.: Erweiterte Wirtschaftlichkeitsanalyse mithilfe des Analytic Hierarchy Process (AHP) unter Berücksichtigung des Wissensmanagements zur Beurteilung von Filialen eines Handelsunternehmens. Universität Duisburg-Essen (Campus Essen), Essen 2006.
- Nr. 31: Zelewski, S.: Beurteilung betriebswirtschaftlichen Fortschritts – ein metatheoretischer Ansatz auf Basis des „non statement view“ (Langfassung). Universität Duisburg-Essen (Campus Essen), Essen 2006.
- Nr. 32: Kijewski, F.; Moog, M.; Niehammer, M.; Schmidt, H.; Schröder, K.: Gestaltung eines Vorgehensmodells für die Durchführung eines Promotionsprojekts am Fachbereich Wirtschaftswissenschaften der Universität Duisburg-Essen, Campus Essen, zum Erwerb des „Dr. rer. pol.“ mithilfe von PETRI-Netzen. Universität Duisburg-Essen (Campus Essen), Essen 2006.
- Nr. 33: Peters, M. L.; Zelewski, S.: Effizienzanalyse unter Berücksichtigung von Satisfizierungsgrenzen für Outputs – Die Effizienz-Analysetechnik EATWOS. Universität Duisburg-Essen (Campus Essen), Essen 2006.

- Nr. 34: Häselhoff, I.; Meves, Y.; Munsch, D.; Munsch, S.; Schulte-Euler, D.; Thorant, C.: Anforderung an eine verbesserte Lehrqualität – Qualitätsplanung mittels House of Quality. Universität Duisburg-Essen (Campus Essen), Essen 2007.
- Nr. 35: Zelewski, S.: Das ADL-Modell der Prinzipal-Agent-Theorie für die Just-in-Time-Produktionssteuerung – Darstellung, Analyse und Kritik. Universität Duisburg-Essen (Campus Essen), Essen 2008.
- Nr. 36: Peters, M. L.; Zelewski, S.: Analyse der Effizienzentwicklung von Bankfilialen mithilfe des Operational Competitiveness Ratings (OCRA). Universität Duisburg-Essen (Campus Essen), Essen 2010.
- Nr. 37: Peters, M. L.; Zelewski, S.: Fallstudie zu Porters generischen Wettbewerbsstrategien im Kontext nachhaltigen Wirtschaftens. Universität Duisburg-Essen (Campus Essen), Essen 2010.
- Nr. 38: Peters, M. L.; Zelewski, S.: Erweiterung von EATWOS um die Berücksichtigung von Satisfizierungsgrenzen für Inputs. Universität Duisburg-Essen (Campus Essen), Essen 2012.
- Nr. 39: Bergenrodt, D.; Jene, S.; Zelewski, S.: Implementierung des Tau-Werts. Universität Duisburg-Essen (Campus Essen), Essen 2013.
- Nr. 40: Millan-Torres, J.; Arndt, C.: Erstellung eines Businessplans zur Existenzgründung des Unternehmens Cowdy! – Anwendung des „Fast-Casual“-Konzepts auf ein systemgastronomisch organisiertes Restaurant mit dem Schwerpunkt der Steakzubereitung. Universität Duisburg-Essen (Campus Essen), Essen 2014.
- Nr. 41: Klumpp, M.; Oeben, M.; Zelewski, S.: Evaluation internationaler Bildungstransfer – Konzeptioneller Rahmen und Diskurs zur wissenschaftlichen Bewertung im Forschungs- und Transferprojekt OpporTUNItY. Universität Duisburg-Essen (Campus Essen), Essen 2018.
- Nr. 42: Oeben, M.; Gerlach, A.-T.; Akdogan, D.; Arabaci, T.; Bagbasi, F.; Gudieva, A.; Klumpp, M.: Evaluation von Bildungsleistungen in Deutschland und Tunesien – das Beispiel des Hochschulsektors. Universität Duisburg-Essen (Campus Essen), Essen 2018.
- Nr. 43: Oeben, M.; Klumpp, M.: Die Berufsschulsysteme in Tunesien und Deutschland – Ein systematischer Vergleich im Rahmen der wissenschaftlichen Evaluation des Projektes OpporTUNItY. Universität Duisburg-Essen (Campus Essen), Essen 2018.
- Nr. 44: Peters, M. L.; Zelewski, S.: Adaption der Efficiency Analysis Technique With Input and Output Satisficing (EATWIOS) zur Berücksichtigung von unteren und oberen Satisfizierungsgrenzen. Universität Duisburg-Essen (Campus Essen), Essen 2018.
- Nr. 45: Oeben, M.; Klumpp, M.: Export von Expertise im Bereich der Berufsausbildung – Erfolgsfaktoren und Hemmnisse für den Aufbau und Betrieb eines technischen Berufsschulzentrums in Tunesien im Forschungs- und Transferprojekt OpporTUNItY. Universität Duisburg-Essen (Campus Essen), Essen 2019.
- Nr. 46: Oeben, M.; Klumpp, M.; Zelewski, S.: Internationaler Bildungstransfer – Internationaler Quervergleich als komparativer Ansatz zu Erfahrungen im Bildungstransfer in Richtung Tunesien. Universität Duisburg-Essen (Campus Essen), Essen 2019.

- Nr. 47: Schagen, J. P.; Zelewski, S.; Heeb, T.: Erhebung und Analyse der Anforderungen an ein KI-Tool aus der Perspektive der betrieblichen Praxis – mit Fokus auf der Wiederverwendung von Erfahrungswissen im Bereich des betrieblichen Projektmanagements. Zugleich KI-LiveS-Projektbericht Nr. 1. Universität Duisburg-Essen (Campus Essen), Essen 2020.
- Nr. 48: Schagen, J. P.; Zelewski, S.; Haselhoff, T.; Schmitz, S.; Heeb, T: Überblick über potenzielle Quellen für Test- und Evaluierungsdaten eines KI-Labors im Rahmen des KI-LiveS-Projekts. Zugleich KI-LiveS-Projektbericht Nr. 2. Universität Duisburg-Essen (Campus Essen), Essen 2021.