KOMPETENZZENTRUM
WasserBerlin

Digitalbevaring.dk

Schlussbericht

Einzelvorhaben FAKIN
Forschungsdatenmanagement an kleinen Instituten

Zuwendungsempfänger:
Kompetenzzentrum Wasser Berlin gGmbH

Förderkennzeichen: 16FDM007

Bundesministerium
für Bildung
und Forschung

Fördermaßnahme:	Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen
Einzelvorhaben:	FAKIN: Forschungsdatenmanagement an kleinen Instituten (Kompetenzzentrum Wasser Berlin gGmbH)
Förderkennzeichen	16FDM007
Laufzeit:	01.05.2017 bis 31.04.2019, kostenneutral verlängert bis 31.07.2019
Selbstkosten:	157.723,16 € (Nachkalkulation)
Zuwendung:	126.132,80 € (80% Förderquote)
Kontakt:	Kompetenzzentrum Wasser Berlin Cicerostr. 24, 10709 Berlin Michael Rustler Tel.: +49 30 53653 825 E-Mail: michael.rustler@kompetenz-wasser.de www.kompetenz-wasser.de
Autoren dieses Berichtes:	Michael Rustler Hella Schwarzmüller

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

INHALTSVERZEICHNIS

Abkürzungen und Glossar 4
I. Kurze Darstellung 5
I.1 Aufgabenstellung 5
I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde 5
I.3 Planung und Ablauf des Vorhabens 6
I.4 Anknüpfung an den wissenschaftlichen und technischen Stand 8
I.5 Zusammenarbeit mit anderen Stellen 9
II. Eingehende Darstellung 10
II.1 Verwendung der Zuwendung und erzielte Ergebnisse 10
Arbeitspaket 1 10
Aufnahme und Analyse von Best Practices (AP 1.1) 10
Recherche und Bewertung von neuen Werkzeugen (AP 1.2) 12
Arbeitspaket 2 13
Anwendung der Best-practices und neuen Werkzeuge für zwei Beispielprojekte (Task 2.1) 13
Implementierung ins QMS (Task 2.2) 16
Arbeitspaket 3 17
Workshops (AP 3.1) 17
Erfahrungsaustausch, Schulungen & Wissensskalierung (AP 3.3) 20
II.2 Die wichtigsten Positionen des zahlenmäßigen Nachweises 25
II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit 25
II.4 Darstellung des voraussichtlichen Nutzens 25
II.5 Während der Durchführung des Vorhabens bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen 26
II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse 27
III. Anhang 28
IV. Literatur 42

ABKÜRZUNGEN UND GLOSSAR

EVA-Prinzip: Prinzip der Informatik, das strikt zwischen **E**ingabe, **V**erarbeitung und **A**usgabe trennt. Dieses Prinzip lässt sich auch unabhängig von Programmiertätigkeiten anwenden, z.B. beim Arbeiten mit einem Tabellenkalkulationsprogramm. Es verlangt dann, dass die Eingangsdaten (Eingabe) in einer eigenen Datei gehalten werden und die Verarbeitungsregeln (Verarbeitung) in einer anderen Datei, die auf die Datei mit den Eingangsdaten verweist. Ebenso sollten Darstellungen (Ausgabe) in eigenen Dateien gehalten werden, die auf die Dateien, in denen die Daten verarbeitet werden, verweisen. Beim Arbeiten mit Datenbankmanagementsystemen ist entsprechend darauf zu achten, zwischen der Datenschicht (Backend = Eingabe) und der Ansichten-Schicht (Frontend = Ausgabe) zu unterscheiden.

FDM: Forschungsdatenmanagement

KWB: Kompetenzzentrum Wasser Berlin

LCA: Life Cycle Assessment (Lebenszyklusanalyse)

Programmcode: "werden die Anweisungen bezeichnet, die im Rahmen der Softwareentwicklung für ein bestimmtes Computerprogramm oder einen Teil davon entstehen und die dessen Funktionalität in einer bestimmten Programmiersprache beschreiben bzw. repräsentieren." (Wikipedia 2020a)

QMS: Qualitätsmanagementsystem

Repositorium: "ist ein verwaltetes Verzeichnis zur Speicherung und Beschreibung von digitalen Objekten für ein digitales Archiv. Bei den verwalteten Objekten kann es sich beispielsweise um Programme (Software-Repository), Publikationen (Dokumentenserver), Datenmodelle (Metadaten-Repository) oder betriebswirtschaftliche Verfahren handeln. Häufig beinhaltet ein Repository auch Funktionen zur Versionsverwaltung der verwalteten Objekte." (Wikipedia 2020b)

I. KURZE DARSTELLUNG

I.1 Aufgabenstellung

Zum Forschungsdatenmanagement zählen alle Aktivitäten, die mit der Aufbereitung, Speicherung, Archivierung und Veröffentlichung von Forschungsdaten verbunden sind. Die Bedeutung des Forschungsdatenmanagements ist in den vergangenen Jahren immens gestiegen. Grund dafür sind die großen Datenmengen, die im Zuge der Digitalisierung und Automatisierung von Prozessen anfallen und neue Herausforderungen an deren Verwaltung und Verarbeitung stellen, die mit den bisherigen Werkzeugen schwer bewältigt werden können. Dies gilt auch für Daten in der Wasserforschung. Der nachhaltige Zugang zu Forschungsdaten und die Erstellung von Datenmanagementplänen werden zunehmend von Forschungsförderern verlangt.

Am Kompetenzzentrum Wasser Berlin gGmbH (KWB) werden im Rahmen von Forschungsprojekten eine Vielzahl von Daten verarbeitet, die entweder selbst erhoben oder von Projektpartnern zur Verfügung gestellt werden. Dazu zählen Messdaten, Metadaten, Fotos/Videos, Bestands- und Zustandsdaten und verarbeitete Daten (z.B. Zeitreihen, aggregierte Werte, Ergebnisse aus Computersimulationen). Um solche Daten nachhaltig nutzbar zu machen, zu verwalten und zu verarbeiten, sind standardisierte Prozesse, Werkzeuge und Methoden zu entwickeln, die eine projektübergreifende Reproduzierbarkeit der Ergebnisse gewährleisten.

Ziel des Projektes FAKIN (Forschungsdatenmanagement an kleinen Instituten) war es, ein solches Forschungsdatenmanagement (FDM) für das KWB in Zusammenarbeit mit den Projektwissenschaftlern zu erarbeiten und unternehmensweit zu etablieren. Damit sollte das Vorhaben als übertragbares Fallbeispiel für das FDM an kleinen, aber stark vernetzten außeruniversitären Forschungsinstituten dienen.

I.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

Das Vorhaben FAKIN wurde im Rahmen der BMBF-Fördermaßnahme „Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen“ durchgeführt.

Das vorliegende Projekt FAKIN sollte das Wissen in folgenden Punkten erweitern:

- Übertragung von FDM auf ein kleines Institut (< 100 Mitarbeiter, keine eigene IT Abteilung) am Beispiel der Kompetenzzentrums Wasser Berlin gGmbH
- Entwicklung von „Best-practices“ und Testen von Werkzeugen zum FDM für kleine Institute
- Entwicklung eigener FDM Tools (z.B. Speicher-, Pfad-, Namensanalyse, Wissensspeicher) zur Unterstützung der Mitarbeiter
- Integration des FDM in Qualitätsmanagementsystem des KWB
- Wissenstransfer an ähnlich kleine außeruniversitäre Forschungsinstitute (Abschlussworkshop)

I.3 Planung und Ablauf des Vorhabens

Das Vorhaben FAKIN startete am 01.05.2017. Eine erste Veranstaltung des BMBF mit dem Ziel der Vernetzung aller geförderten Projekte aus der BMBF-Fördermaßnahme „Erforschung des Managements von Forschungsdaten in ihrem Lebenszyklus an Hochschulen und außeruniversitären Forschungseinrichtungen“ fand am 1. Juni 2017 in Berlin statt.

Die Durchführung des Forschungsvorhabens folgte den im Projektantrag beschriebenen Arbeitspaketen (Abbildung 1). Die Bearbeitung der einzelnen Arbeitspakete wurde in zwei Zwischenberichten dokumentiert. Eine Übersicht der Veranstaltungen des Projektes ist in Tabelle 1 zusammengestellt.

Die Arbeiten am KWB starteten mit Arbeitspaket 1 „Identifikation“ und einem internen Auftaktworkshop im September 2017 (Task 3.1.1.) zur Identifikation und Priorisierung der an unserem kleinen Institut wichtigsten Forschungsdatenmanagement (FDM) -Themen. Für diese wurden im nächsten Schritt sowohl „Best-practices“ entworfen (Task 1.1) als auch neue Werkzeuge (Task 1.2) evaluiert. Die Ergebnisse wurden in einem zweiten internen „Best-practices“ Workshop im Januar 2018 den Kollegen vorgestellt und mit ihnen diskutiert.

Anschließend erfolgte im Arbeitspaket 2 „Implementierung“ eine zirka einjährige Testphase (Task 2.1) der verabschiedeten „Best-practices“ anhand von zwei ausgewählten Forschungsprojekten des KWB. Die daraus resultierenden Erfahrungen wurden in einem dritten internen „Lessons-learnt“ Workshop (Task 3.1.3) im März 2019 noch einmal Kollegen aus allen drei Forschungsabteilungen vorgestellt und mit diesen abgestimmt, welche Prozesse in das um Datenprozesse zu erweiternde Qualitätsmanagementsystem des KWB (Task 2.2) aufgenommen werden sollten.

Das Arbeitspaket 3 „Kommunikation“ beinhaltete neben den oben angeführten internen Workshops (Task 3.1.1 – 3.1.3) auch noch einen Abschlussworkshop mit externen Instituten (Task 3.1.4), der am 25. Juli 2019 stattfand. Darüber hinaus wurde projektbegleitend die Task 3.3 „Erfahrungsaustausch, Schulungen & Wissensskalierung“ bearbeitet. Daraus entstanden Schulungsunterlagen und Tutorials für die interne Mitarbeiterschulung, Publikationen (z.B. Software-Algorithmen in öffentlichen Repositorien, siehe: https://zenodo.com/communities/kwb) und ein Prototyp für ein institutionelles Wissensrepositorium.

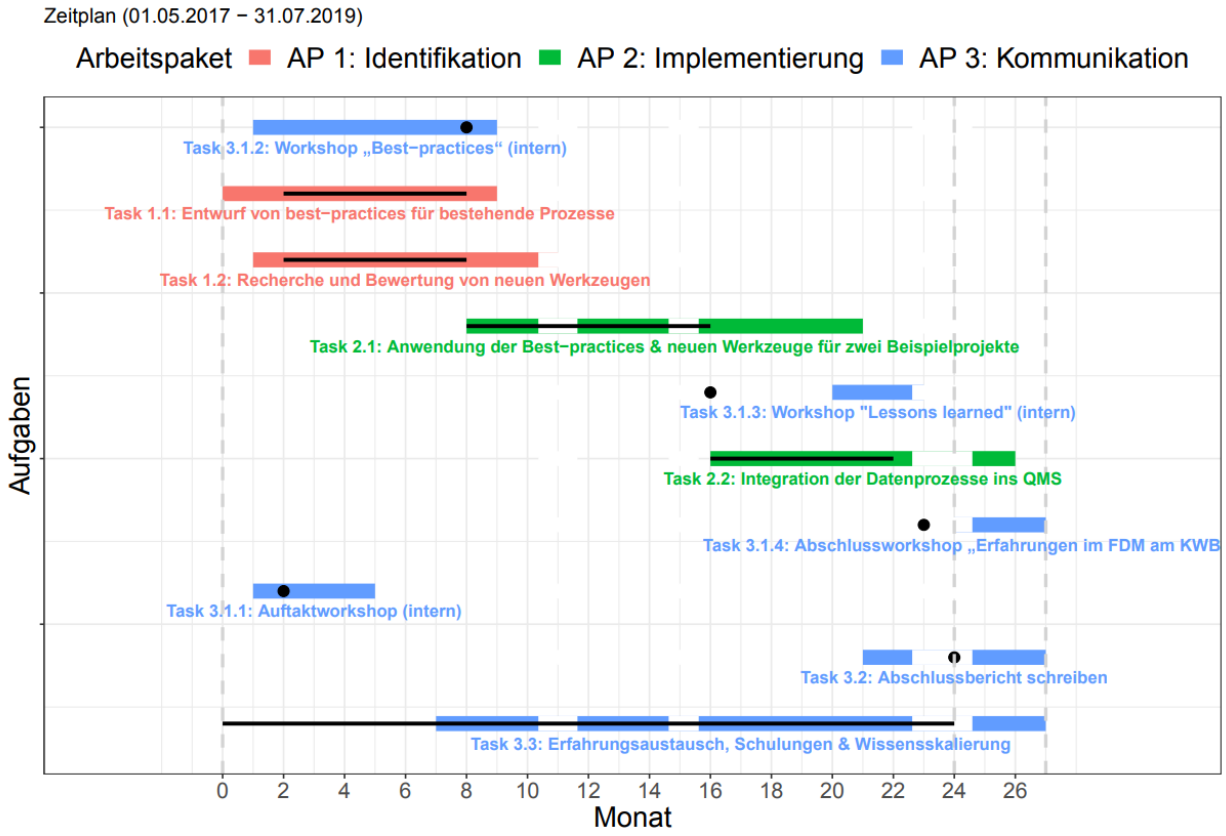


Abbildung 1 **Gantt-Chart des Projektes FAKIN (schwarze Linien/Punkte: ursprünglicher Zeitplan laut Projektantrag)**

Tabelle 1: Veranstaltungen des Projektes FAKIN

Datum	Ort	Kategorie	Bezeichnung
01.06.2017	Berlin	Präsentation (extern)	FDM-Netzwerkveranstaltung des BMBF
07.07.2017	Berlin	Präsentation (intern)	Was ist FAKIN? Was ist FDM?
21.09.2017	Berlin	Workshop (intern)	Auftaktworkshop
20.01.2018	Berlin	Workshop (intern)	Best-practices Workshop
21.09.2018	Berlin	Präsentation (intern)	Zwischenergebnisse: FAKIN Testprojekte
19.10.2018	Berlin	Präsentation (extern)	FDM-Netzwerkveranstaltung des BMBF
16.02.2019	Berlin	Präsentation (extern)	Knowledge Repo: An Innovative Way for Sharing Knowledge At An Institutional Level (Session auf Open Science Barcamp)
06.03.2019	Berlin	Workshop (intern)	Lessons-learnt Workshop
24.07.2019	Berlin	Workshop (extern)	Abschlussworkshop FAKIN
20.08.2019	Marburg	Präsentation (extern)	Wie etablieren wir FDM an unserem kleinen Wasserforschungsinstitut? (Workshop: FDM praktikabel gestalten, Herder Institut)

I.4 Anknüpfung an den wissenschaftlichen und technischen Stand

Forschungsdatenmanagement umfasst den gesamten Datenzyklus von der Datenerhebung zur Publikation und endet bei der Archivierung und Nachnutzbarmachung (Abbildung 2).

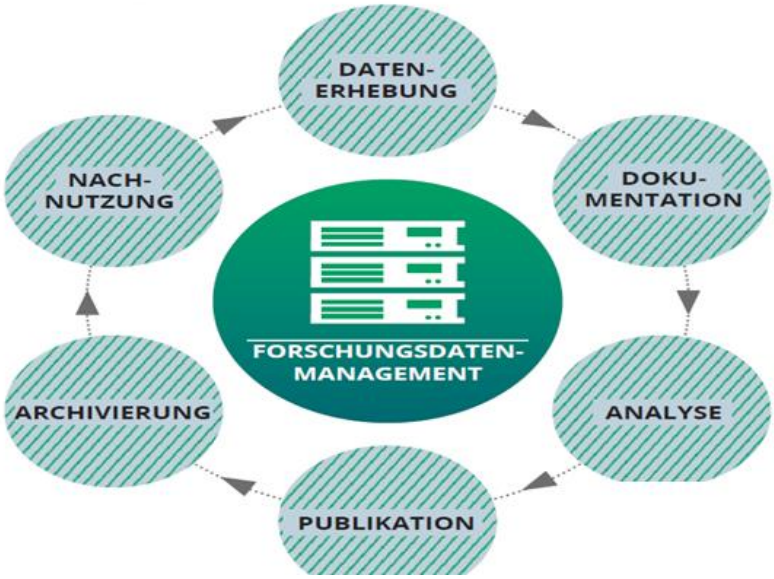


Abbildung 2 Forschungsdatenzyklus (aus: Bertelmann et al. 2014)

Am KWB ist das Forschungsdatenmanagement ein relativ neues Themenfeld, dessen Wichtigkeit jedoch nicht nur durch die gestiegenen Anforderungen bei der Projektentwicklung (z.B. Datenmanagementpläne verpflichtend bei EU Horizon 2020 Open Data Research Pilots, (EC 2017)) in den letzten Jahren immer stärker zunimmt. Ein weiterer Grund ist, dass bei der Datenerhebung an unserem Institut immer häufiger Datenlogger eingesetzt werden, die eine große Menge Rohdaten generieren. Diese lassen sich mit konventionellen Analyse-Werkzeugen (z.B. MS EXCEL) aufgrund der Datenmenge (> 1 Mio. Datenpunkte) nicht mehr analysieren, so dass eine automatisierte Datenverarbeitung mit Hilfe von Programmierung unter Nutzung von Versionsverwaltungssoftware (Git 2020; Subversion 2020) eingesetzt wird (Sonnenberg et al. 2013) und Repositorien (GitHub 2020; Zenodo 2020) zur Veröffentlichung von Programmcode (Rustler 2016a; Rustler 2016b; Sonnenberg 2016).

Ziel in FAKIN war es daher, die oben genannten Vorarbeiten einzelner Mitarbeiter am KWB zu den Themenkomplexen Datenerhebung, Analyse, und Publikation auf den gesamten Forschungsdatenzyklus auszuweiten (d.h. Dokumentation, Archivierung, Nachnutzung) und zudem auf institutioneller Ebene zu verankern. Hierzu soll das bereits seit dem Jahr 2010 am KWB existierende Qualitätsmanagementsystem (KWB 2019) – das beispielsweise Prozesse für das Projektmanagement enthält – um einen Punkt zum FDM ergänzt werden.

Die Etablierung des FDM auf institutioneller Ebene hat mit der Erstellung von ersten Forschungsdatenpolicies an deutschen Universitäten ungefähr im Jahr 2013 begonnen (Helbig et al. 2019; Hiemenz & Kuberek 2018b). Bis vor kurzen existierte zu diesem Thema kaum praxisnahe Literatur. Diese Wissenslücke wurde jedoch im „FDMentor“ Projekt geschlossen, da hier mehrere Berichte entstanden sind (FDMentor 2019), die sich zwar primär an Universitäten und größere Forschungseinrichtungen richten, jedoch Teilaspekte auch an kleinen Forschungsinstituten anwendbar sind. Beispielsweise wurde ein Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (Hartmann et al. 2019) entwickelt, das einen strukturierten Bewertungsrahmen zur Selbstevaluation und Zielbestimmung bietet und sich daher als Werkzeug zur FDM-Strategieentwicklung an Forschungseinrichtungen eignet. Für die Erstellung einer institutionellen Forschungsdatenpolicy empfiehlt sich die Nutzung des „Forschungsdaten-Policy-Kits“ (Hiemenz & Kuberek 2018a), der als generischer Baukasten mit Leitfragen und Textbausteinen ebenfalls im FDMentor Projekt entwickelt wurde.

I.5 Zusammenarbeit mit anderen Stellen

Neben dem intensiven Austausch mit den Mitarbeitern des KWB im Rahmen der Workshops und der Arbeit an den Beispielprojekten, und dem ständigen Austausch mit Projektpartnern im Rahmen der Forschungsprojekte des KWB, wurden in FAKIN zwei Projekte anderer Stellen identifiziert und eine Zusammenarbeit initiiert. Darüber hinaus fand ein Abschlussworkshop mit externer Beteiligung statt und das KWB wurde angefragt, das Projekt FAKIN bei einer Veranstaltung des Herder-Institutes zu präsentieren.

DFG Projekt „o2r“ (opening reproducible research, https://o2r.info) der Universität Münster

Es wurde im Rahmen von FAKIN im Oktober 2018 getestet, ob sich in der Programmiersprache R (R Core Team 2020) berechnete Modellierungsergebnisse in einem bereits abgeschlossenen KWB-Forschungsprojekt mit Hilfe des o2r Tools “executable research compendium” (https://o2r.uni-muenster.de) reproduzieren lassen. Dies ist bis auf geringfügige Abweichungen (z.B. unterschiedliche Farbgebung in Abbildungen) gelungen, wie hier gezeigt: https://o2r.uni-muenster.de/#!/erc/Bcrvh. Dieser Reproduzierbarkeitstest stellt den ersten erfolgreichen Anwendungsfall außerhalb des o2r Projektkonsortiums dar.

Metadatenstandard „codemeta“

Das "codemeta"-Projekt definiert ein "JSON-LD"-Format zur Beschreibung von Software-Metadaten (CodeMeta 2019).

An unserem Institut spielt die Programmierung, vor allem in der Programmiersprache R (R Core Team 2020), eine große Rolle. Beispielsweise hat unser Institut mittlerweile ca. 40 R Pakete (https://kwb-r.github.io/status) veröffentlicht. Ein Tool zur Erstellung einer „codemeta“ Metadatendatei haben Carl Boettiger (UC Berkeley) und Maelle Salmon (rOpenSci, https://ropensci.org) mit dem R Paket „codemetar“ entwickelt (Boettiger et al. 2019). Durch Codebeiträge von Hauke Sonnenberg und Michael Rustler konnte dieses Werkzeug so verbessert werden, dass darauf basierend für unser Institut im Rahmen von FAKIN eine Routine entwickelt werden konnte: Einmal täglich wird nun mit Hilfe des R Paketes „pkgmeta“ (Rustler 2020b) automatisiert eine solche „codemeta“ Metadatendatei für alle öffentlichen R-Pakete unseres Instituts erstellt und unter https://kwb-r.github.io/pkgmeta/codemetar.json abgelegt. Damit lassen sich erstmals detaillierte Analysen zu Softwareabhängigkeiten durchführen, wodurch sich in Zukunft die potentiellen Auswirkungen von Änderungen in einer Codebasis besser einschätzen lassen (https://kwb-r.github.io/pkgmeta/articles/codemetar-analysis.html).

Wissenstransfer

Am 25. Juli 2019 wurden die Ergebnisse und Erfahrungen aus dem FAKIN Projekt 14 Teilnehmern ähnlich kleiner Forschungsinstitute oder Ingenieurbüros vorgestellt. Darüber hinaus konnten die Projektergebnisse auch auf dem Workshop „Forschungsdatenmanagement praktikabel gestalten“ des Herder Instituts in Marburg am 20. August 2019 vorgestellt werden.

Auf beiden Veranstaltungen wurde großes Interesse an einem innerhalb von FAKIN entwickeltem Tool zur Pfadanalyse der eigenen Datenstrukturen bekundet, so dass aufgrund des positiven Feedbacks beschlossen wurde, dieses nach Projektabschluss unter einer open-source Lizenz (MIT) zu veröffentlichen und damit der FDM-Community zur Verfügung zu stellen (Sonnenberg & Rustler 2020).

II. EINGEHENDE DARSTELLUNG

II.1 Verwendung der Zuwendung und erzielte Ergebnisse

Das wissenschaftliche Ziel des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Forschungsprojektes FAKIN (Forschungsdatenmanagement an kleinen Instituten) war die Entwicklung standardisierter Prozesse, Instrumente und Methoden zur Reproduzierbarkeit von Forschungsergebnissen.

In den folgenden Abschnitten werden die Ergebnisse, nach Arbeitspaketen sortiert, vorgestellt.

Arbeitspaket 1

Aufnahme und Analyse von Best Practices (AP 1.1)

Um festzulegen, für welche konkreten Arbeitsabläufe das Projekt FAKIN Best Practices erarbeiten sollte, wurde zunächst der Ist-Zustand bezüglich des Umgangs mit Daten am KWB aufgenommen. Dies erfolgte im Auftaktworkshop. Um die individuellen Problemstellungen und Bedarfe zu den im Auftaktworkshop ermittelten Schwerpunktthemen weiter zu konkretisieren, wurden Interviews mit insgesamt sechs Projektleitern aus allen drei Fachabteilungen des KWB durchgeführt. Die Interviews wurden jeweils einzeln schriftlich zusammengefasst. Sie lieferten Informationen darüber, mit welchen Arten und Größenordnungen von Daten in den unterschiedlichen Projekten umgegangen wird, wie Daten von Projektpartnern erhalten, wo und in welcher Form sie abgelegt und wie sie üblicherweise verarbeitet werden. Die Mitarbeiter berichteten auch darüber, was bei ihnen gut funktioniert, in welchen Bereichen es häufiger zu Schwierigkeiten kommt, wo sie Potential für Verbesserungen sehen und wofür sie sich Lösungen durch FAKIN wünschten. Die in den Interviews konkret genannten Problemfelder des Datenmanagements und deren komplexe Abhängigkeiten wurden in einer Mind-Map grafisch dargestellt (Anhang A). Mit Hilfe der Mind-Map konnten Schwerpunktthemen identifiziert werden, die für viele Mitarbeiter relevant sind und daher prioritär in FAKIN bearbeitet wurden: *Rohdaten*, *Metadaten*, *Versionen* von Daten/Dateien, *Ordnerstruktur* und *Nomenklatur* (Ordner- und Dateinamen). Der Umgang mit diesen Themen war innerhalb des KWB uneinheitlich und sollte im Rahmen von FAKIN vereinheitlicht werden. Im Folgenden werden die erarbeiteten Empfehlungen detailliert beschrieben.

Ordnerstrukturen

Die Projektdateiablagestruktur an unserem Forschungsinstitut wird durch die Anwendung des EVA Prinzips aus der Informatik, d.h. die Trennung von Eingabe (Rohdaten), Verarbeitung (Datenverarbeitung) und Ausgabe (Ergebnisse), verbessert. Diese Trennung wird im Folgenden näher erläutert.

> ***Rohdaten*** werden auf einem besonders schreibgeschützten Bereich auf unserem Institutsserver (//server/rawdata/project-identifier) abgespeichert und somit vor versehentlichem Löschen geschützt.

> Die ***Datenverarbeitung*** erfolgt in einem gesonderten Bereich auf unserem Institutsserver (//server/processing/project-identifier). Unterordner werden thematisch angelegt (z.B. nanofiltration_tiefwerder). Die für die Datenverarbeitung benötigten Rohdaten werden möglichst automatisiert vom Rohdatenserver eingelesen und weiterverarbeitet. Bei Programmierung ist die Nutzung von Versionsverwaltungssoftware verpflichtend, ansonsten wird eine manuelle Versionierung empfohlen. Eine automatisierte Verarbeitung wird empfohlen, wenn Daten aus vielen gleichartigen Dateien (z.B. Datenlogger) zusammengeführt werden. Die Datenverarbeitungsebene kann als „Spielwiese“ gesehen werden, auf der große Datenmengen beim Testen von verschiedenen Varianten (z.B. Szenarien Analysen, Modellvarianten) anfallen können, von denen jedoch nur einige wenige „berichtsrelevant“ werden.

Ergebnisse die auf der Datenverarbeitungsebene erstellt wurden und berichtsrelevant sind, sind mit einer Versionsnummer zu versehen und sind durch Verlinkung von der Projektergebnisebene (//server/results/project-identifier) auf die Datenverarbeitungsebene auffindbar.

Bei der Projektergebnisebene handelt sich um eine administrative Projektleitersicht, in der die Ordnerbenennung nicht thematisch sondern nach Arbeitspaketen oder Berichten erfolgt. Bei Integration von verarbeiteten Daten (z.B. Tabellen, Abbildungen) in Berichten sind Links zur Datenverarbeitungsebene zu setzen. Vorteil dieses Vorgehens ist, dass durch Auslagerung der Datenverarbeitung auf ein eigenes Serververzeichnis, der administrative Projektordner übersichtlich gehalten werden kann und zugleich die Nachvollziehbarkeit hinsichtlich der Ergebnis- und Berichtsversionen verbessert wird.

Darüber hinaus sind die folgenden allgemeinen Regeln zur Orderstruktur und Dateibenennung festgelegt worden ((Rustler et al. 2019a), https://kwb-r.github.io/fakin.doc/best-practices.html#rule-d-avoid-long-names):

- Ordnerstukturen nicht zu tief
- Ordner- (< 20 Zeichen) und
- Dateinamen (< 50 Zeichen) nicht zu lang

Grund für diese Beschränkung ist, dass auf Windows Betriebssystemen die maximale Pfadlänge auf 260 Zeichen beschränkt ist. Bei Nichtbeachtung der oben genannten Regeln erfolgt das Kopieren oder Verschieben von Ordnerstrukturen in „tiefere Ordnerebenen“ (z.B. von „\\servername\exchange\user“ nach „\\servername\department\grw\project\fakin\wp1\“) gegebenenfalls nur unvollständig, falls die maximale Pfadlänge überschritten ist.

Metadaten

Es wird ein dezentraler Ansatz bevorzugt, d.h. Metadaten stehen bei den Daten anstatt in einer zentralen Metadatenbank. Es ist besser, unstrukturierte oder schlecht formulierte Metadaten als keine Metadaten zu erheben. Allerdings ist es schwierig, einen passenden Standard zu finden. Daher wird folgende Minimalanforderung definiert: Es sind einfache Textdateien (z.B. README.txt) zu erstellen, die die im Ordner enthaltenen Ordner oder Dateien beschreiben und direkt bei den Daten abgelegt werden.

Eine Erweiterung dieses Ansatzes ist die Erstellung von strukturierten Textdateien im yaml-Format, wie für Projektmetadaten in folgendem Beispiel in Abbildung 3 gezeigt wird („projects.yaml“):

PROJECTS.yml

```yaml
aquanes:
  department: wwt
  short-name: AquaNES
  long-name: >
    Combined Processes for Water Treatment Systems
  financing:
    funder: EU-H2020

fakin:
  department: grw
  short-name: FAKIN
  long-name: >
    Forschungsdatenmanagement an kleinen Instituten
  financing:
    funder: BMBF
```

Abbildung 3 **Beispiel eine strukturierten Metadatendatei (YAML Format) für Projekte**

Nomenklatur

Es werden keine Umlaute, Leer- oder Sonderzeichen in Ordner- oder Dateinamen zugelassen. Der Bindestrich ist zu verwenden, um zusammengesetzte Wörter zu trennen, der Unterstrich, um unterschiedliche Informationen abzugrenzen.

Darüber hinaus wird eine einheitliche Datumsschreibweise nach dem Schema „yyyy-mm-dd“ (ISO-8601 2019) festgelegt. Es wird empfohlen einfache, aber "sprechende" Namen zu verwenden.

Dateinamen können relativ lang sein (< 50 Zeichen), sollten aber im Idealfall einer Nomenklatur folgen, die projektbezogen festgelegt werden sollte.

Versionierung

Die Nutzung von Versionsverwaltungssoftware (Subversion, Git) ist obligatorisch für die Verwaltung von Programmcode. Wer Skripte programmiert, verwendet unsere interne Versionsverwaltung Subversion. Wer R-Pakete entwickelt, verwendet Git und veröffentlicht diese auf der Codeplattform GitHub. Für letzteres wird die Nutzung des im Rahmen von FAKIN erstellten Hilfspaketes kwb.pkgbuild (Rustler & Sonnenberg 2019) empfohlen, da dieses eine schnelle Paketgrundgerüsterstellung in hoher Qualität im Corporate Design ermöglicht.

Eine manuelle Versionierung mit dreiteiliger Versionsnummer wird bei Berichten oder Daten empfohlen. Beispiel: alles mit 0.x.y ist "draft", alles ab 1.0.0 ist "final".

Besondere Herausforderungen

An unserem Institut werden Informationen von Partnern häufig in Form von Tabellen (MS Excel) zur Verfügung gestellt, die für die automatisierte Datenverarbeitung eine besondere Herausforderung darstellen, da sie i.d.R. erst nach einer zeitintensiven manuellen Datenaufbereitung weiterverarbeitet werden können. Zur Erreichung einer hohen Datenqualität von Tabellen sind folgende Regeln in Anlehnung an DataCarpentry (2019) und zu beachten:

- Keine Formatierung, keine Leerzeilen oder Leerspalten aus "optischen" Gründen
- Genau eine Zeile mit Spaltennamen, mehrzeilige Header in einzeilige Header überführen, dabei die Zusatzinformationen als Metadaten in eigenes Dokument oder Tabellenblatt schreiben.
- Einfache Spaltennamen ohne Leer- oder Sonderzeichen.
- Ggf. existierendes Datenmodell (z.B. ODM-2 (2020)) nutzen, das Spaltennamen vorgibt

Darüber hinaus ist zur Eindeutigkeit von Zeitinformationen neben der Uhrzeit auch immer der UTC-Offset anzugeben (ISO-8601 2019). Aufgrund der Wichtigkeit dieses Themenkomplexes für das KWB (z.B. Einsatz von Datenlogger), wurde hierzu das R Paket „kwb.datetime“ (Sonnenberg 2019) um ein Tutorial zum Umgang mit Zeitzonen ergänzt (https://kwb-r.github.io/kwb.datetime/articles/timezones.html).

Recherche und Bewertung von neuen Werkzeugen (AP 1.2)

Im Rahmen des Arbeitspakets 1.2 wurden Methoden und Werkzeuge identifiziert, die das Potential haben, das FDM am KWB zu verbessern. Hierbei wurde festgestellt, dass es zu den vielfältigen Problemfeldern des FDM ein noch vielfältigeres und in weiten Teilen unübersichtliches Angebot von frei verfügbaren und kommerziellen Werkzeugen gibt. Es wurde unter anderem zu den Themen Metadatenstandards und Metadateneditoren recherchiert. Es wurde eine Sammlung von Links zu Internetquellen angelegt. Bezüglich neuer Methoden des FDM wurde mit DataOne (2019) eine ausführliche, durchsuchbare Sammlung von Best Practices-Dokumenten identifiziert.

In Bezug auf neue Werkzeuge wurde eine Quelle gefunden, die eine umfangreiche Liste von Werkzeugen (Tools) mit Kurzbeschreibungen zu verschiedenen Themen des Datenmanagements übersichtlich darstellt und Verknüpfungen zu den Anbietern der Werkzeuge liefert (Bosman und Kramer, 2018).

Die Ergebnisse aus AP 1.1 und 1.2 sind ausführlich im internen KWB Bericht „Best-Practices in Research Data Management“ (Rustler et al. 2019a) dargestellt. Zudem wurden die getesteten FDM-Werkzeuge (AP 1.2) im FAKIN Wissensrepositorium (Rustler 2019a) aufgenommen und thematisch verlinkt (https://kwb-r.github.io/fakin/#tool).

Arbeitspaket 2

Anwendung der Best-practices und neuen Werkzeuge für zwei Beispielprojekte (Task 2.1)

Die „Best-practices“ (vgl. AP 1.1) und neuen Werkzeuge wurden für zwei Testprojekte – die im Rahmen des internen Best-Practices Workshops im Januar 2018 festgelegt wurden – über ein Jahr (bis März 2019) auf ihre Praxistauglichkeit geprüft. Nachfolgend sind für beide Projekte die wesentlichen Ergebnisse kurz zusammengefasst.

1) AquaNES (Projektseite, Laufzeit: Juni 2016 – Mai 2019): ist ein EU-Projekt mit 30 Projektpartnern bei dem das KWB zwei Pilotanlagen in Berlin betreibt. Monatlich fallen für jede dieser Anlagen 10 Millionen Betriebsdatenpunkte an, die automatisiert erfasst werden. Die Herausforderung ist die Konzeption einer Datenstruktur die die Auffindbarkeit der Daten verbessert, den Schutz der Rohdaten gewährleistet (Schreibschutz) und zudem auf die Bedürfnisse der automatisierten Datenverarbeitung (Datenimport, -aggregierung, -visualisierung und -export) abgestimmt ist. Eine zweite Herausforderung für das Forschungsdatenmanagement in diesem Projekt stellt die Softwareentwicklung in Zusammenarbeit mit Projektpartnern dar. Es sollte die kollaborative Entwicklung einer Web Applikation zur quantitativen mikrobiellen Risikoeinschätzung unterstützt werden. Im Vordergrund steht hier insbesondere die Entwicklung eines Workflows, um Programcode auf nachvollziehbare und einfache Weise miteinander auszutauschen. Dies umfasst neben der Berechnungsmethodik und deren Umsetzung (Algorithmus) auch die verwendete Datengrundlage (z. B. Zuflusskonzentrationen, Dosis-Wirkungsbeziehungen u. ä.).

Ergebnisse: In diesem Testprojekt konnten exemplarisch die Vorteile einer Best-Practice Ordnerstruktur (Abbildung 4) gezeigt werden (d.h. EVA-Prinzip, Namens- und Ordnerkonventionen). Das Online-Tool Treejack (OptimalWorkshop 2019) diente zur Quantifizierung, ob sich Daten innerhalb der neuen Struktur besser (d.h. schneller und auf direkterem Wege) finden lassen als in der bisherigen Struktur. Hierzu wurden neun Wissenschaftler am KWB mit einer „Datensuche“ beauftragt: es mussten drei Dateien jeweils in der alten und neuen Ordnerstruktur gefunden werden. Die Ergebnisse (Anhang B) zeigten, dass in einer verbesserten Ordnerstruktur diese schneller, direkter (ohne Abbiegen in falsche Unterordner) und deutlich häufiger als in der anfänglichen Ordnerstruktur gefunden wurden.

Allerdings zeigte sich auch, dass Änderungen der Ordnerstruktur in einem laufenden Projekt mit vielen Beteiligten schwierig umzusetzen ist, da beispielsweise Pfadabhängigkeiten und Verknüpfungen bestehen. Die Implementierung der Best-Practices in eine unternehmensweite Strategie ist daher umso wichtiger.

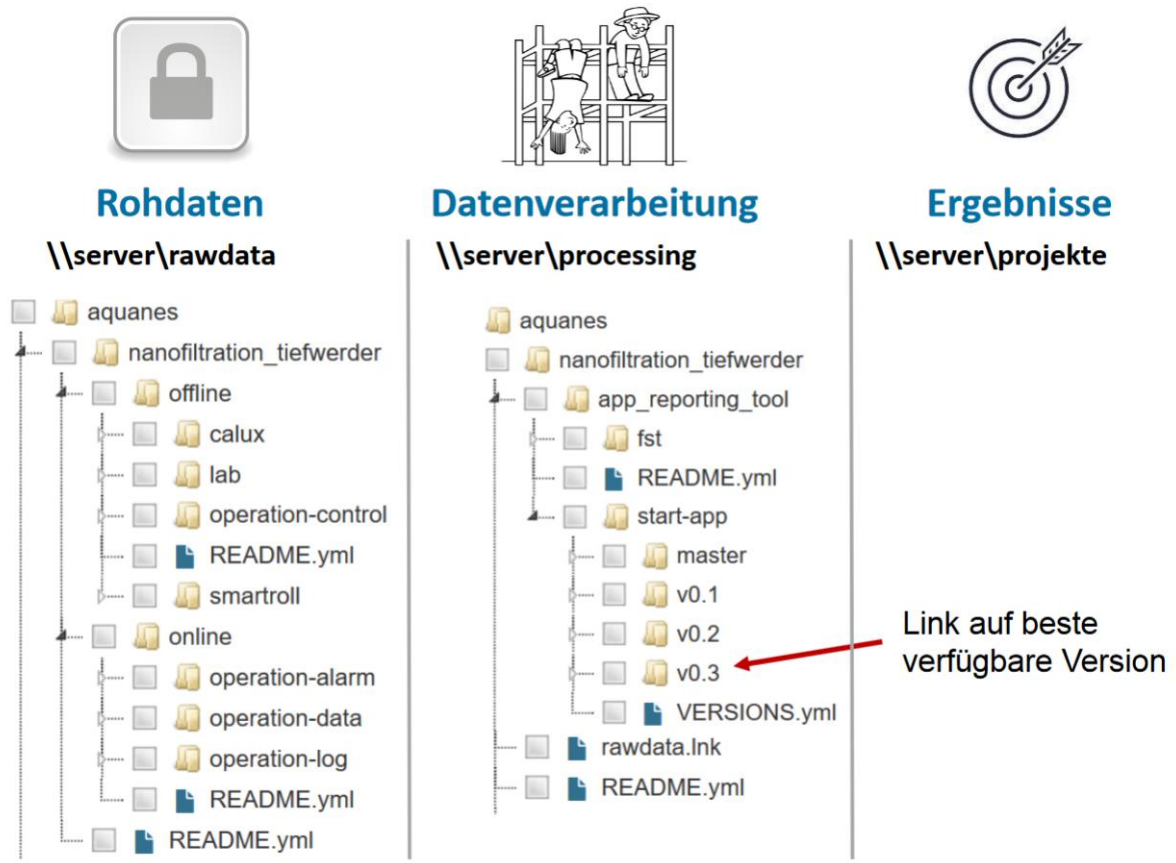


Abbildung 4 **Beispielhafte Umsetzung der „Best-Practices“ für das FAKIN Testprojekt AquaNES**

Für die zweite Herausforderung im Beispielprojekt AquaNES, die Entwicklung von zwei interaktiven Tools zur Auswertung von zeitlich hoch aufgelösten Betriebsdaten von insgesamt vier verschiedenen Pilotanlagen und zur quantitativen mikrobiellen Risikoberechnung, wurde aus FAKIN heraus empfohlen, die Entwicklung der Tools auf der kollaborativen Plattform GitHub (siehe: https://github.com/topics/project-aquanes) vorzunehmen. Beiden gemeinsam war, dass sie:

- externen Partnern zur Verfügung stehen und/oder mit diesen zusammen entwickelt werden
- transparente und nachvollziehbare Methodik/Berechungsalgorithmen und
- als Softwareprojekte koordiniert werden sollten.

Ursprünglich wurde die Microsoft Office 365-Cloudlösung eingesetzt. Die Verlagerung auf GitHub ermöglichte eine automatische Versionierung und Verknüpfung mit dem Repositorium Zenodo zur Langzeitarchivierung. Mit diesem Ansatz konnte der Datenaustausch stark vereinfacht werden, da der Programmcode nicht noch einmal manuell an einen weiteren Ort kopiert werden musste, sondern entweder als ZIP Datei oder mit der Software R direkt von GitHub installiert werden konnte.

Allerdings erwies es sich als schwierig die kollaborativen Möglichkeiten der Plattform GitHub (z.B. Erstellen von „Issues“ z.B. Bug Reports“) vollständig auszunutzen, da dies von den Beteiligten ein hohes Maß an Vorwissen erfordert hätte, das nicht vorausgesetzt werden konnte. Daher ist die Empfehlung aus FAKIN in künftigen Projekten mit den Beteiligten eine ein- bis zweitägige GitHub Schulung zum Projektstart zu planen um dessen kollaborative Möglichkeiten besser auszunutzen.

2) Smart-Plant (Projektseite, Laufzeit: Juni 2016 – Mai 2020)**:** ist ebenfalls ein großes EU-Projekt mit 28 Projektpartnern, bei dem am KWB Modellierungen zum „Life Cycle Assessment“ (LCA) im Vordergrund stehen. Eine Besonderheit von LCA-Modellierungen mit der Simulationssoftware „Umberto“ ist, dass die Eingangsdaten ständig verändert werden und dass sich die Modellergebnisse für unterschiedliche Umberto-Versionen zum Teil erheblich voneinander unterscheiden können. Dies ist in der Tatsache begründet, dass Umberto mit einer Datenbank ausgeliefert wird, die zusätzlich zur Software regelmäßig aktualisiert wird.

Die Herausforderung für das Forschungsdatenmanagement ist die Zuordnung der Eingangsdaten (EXCEL Dateien), der jeweiligen Umberto Version, und dem korrespondierendem Ergebnis herzustellen und die im Abschlussbericht verwendeten Ergebnisse/Visualisierungen mit der jeweiligen Modellkonstellation zu dokumentieren. Daher ist die zusätzliche Erfassung von Metadaten (Softwareversion, Datenbankversion) bei der Erstellung von Abschlussberichten unabdingbar um das Ziel von Reproduzierbarkeit erreichen zu können.

Ergebnisse: Die LCA-Modellierung am KWB erfolgte erst nach Beginn der Testphase. Daher waren noch keine Projektordnerstrukturen entstanden und das (E)VA Prinzip konnte weitgehend angewendet werden. Da bei LCA-Modellierungen jedoch keine „Rohdaten" anfallen, wurden in der Orderstruktur nur die Ebenen *Datenverarbeitung („processing")* und *Ergebnisse („results")* angelegt. Für die Benennung von Ordner- und Dateinamen wurden die in FAKIN definierten „Best-Practices" verwendet. Zusätzlich wurden Metadaten für Ordner / Dateinamen (in Readme Dateien im YAML Format) sowie verwendete Software/Datenbanken (Versionsnummern) entsprechend den FAKIN „Best-Practices (vgl. AP 1.1) angelegt. Des Weiteren erfolgte eine manuelle Versionierung für:

- Umberto-Eingangsdaten: d.h. Fragebögen die an die verschieden Projekt-Partner (ggf. mehrfach) versendet wurden,
- Umberto-Modelldateien und
- Exportierte Modellergebnisse.

Beim Modellergebnisexport aus Umberto handelte es sich um Daten, die noch aggregiert werden mussten. Um Fehler bei der manuellen Datenverarbeitung auszuschließen, wurde auch diese Aufgabe automatisiert. Hierzu wurde das R Paket kwb.umberto (Rustler & Sonnenberg 2020) entwickelt, um damit die Modellergebnisse zu aggregieren und automatisiert in eine EXCEL Datei zu exportieren. Auf diese Datei wird nun aus einer zweiten EXCEL Datei – in dem die Vorlagen zur Erstellung von Graphiken sind – verlinkt, so dass bei Modellaktualisierungen die Abbildungen automatisiert aktualisiert werden. Dieser in FAKIN entwickelte Workflow ermöglicht nun nicht nur eine bessere Reproduzierbarkeit der Forschungsergebnisse, sondern trägt auch dazu bei, dass auch komplexere Modellanwendungen (z.B. Szenarien- und Sensitivitätsanalysen) möglich sind, da der zeitaufwändige und fehleranfällige Schritt der manuellen Datenverarbeitung weitgehend automatisiert wurde.

Implementierung ins QMS (Task 2.2)

Das am KWB bestehende Qualitätsmanagementsystem beinhaltet bereits einen Prozess Projektmanagement und eine Arbeitsanweisung zum Umgang mit Messgeräten. Beide verweisen auf das Erfordernis einer guten und vollständigen Dokumentation aller Arbeitsschritte und Daten, geben allerdings keine Vorgaben zur Vollständigkeit, zu Ordnerstrukturen, Namenskonventionen usw. Das QMS sollte daher basierend auf den FAKIN-Ergebnissen um eine Komponente Forschungsdatenmanagement erweitert werden.

Im Rahmen des internen „Lessons-learnt“ Workshops am 11. März 2019 wurden die Erfahrungen aus der Anwendung der „Best-practices“ (vgl. AP 1.1) innerhalb zweier KWB Testprojekte vorgestellt, diskutiert und mit den Mitarbeitern und der Geschäftsführung des KWB abgestimmt, welche Themen in das unternehmensweite Qualitätsmanagementsystem (QMS) zum Forschungsdatenmanagement aufgenommen werden sollten.

Folgende Inhalte – die detailliert in AP 1.1. erläutert sind – wurden beschlossen:

- **Orderstruktur:** Einrichtung von getrennten Bereichen auf Institutsserver-Ebene für Rohdaten, Datenverarbeitung und Ergebnisse
- **Metadaten:** Definition von Mindestanforderungen (Beschreibung aller Dateien / Unterordner in einem Ordner in einer README Textdatei).
- **Namenskonventionen**: Definition von zulässigen Dateinamen, Projektakronymen, Zeichen
- **Versionierung:** freiwillige, manuelle Versionierung für Dokumente, bei Programmierung jedoch verpflichtende Nutzung von Software (Subversion, Git) zur automatisierten Versionierung. Für kleinere Programme (z.B. R Skripte) wird die Software Subversion und ein im Intranet eingerichteter zentraler Server genutzt. Bei größeren Programmierprojekten (z.B. Entwicklung von R Paketen) wird die Software Git zusammen mit der Online-Plattform GitHub (https://github.com/kwb-r) eingesetzt.
- **Veröffentlichungen:** von Berichten und Programmcode (R Pakete) in öffentlich geförderten Forschungsprojekten erfolgen im Repositorium Zenodo (https://zenodo.org/communities/kwb). Dieses dient nicht nur zur Langzeitarchivierung dar, sondern verbessert darüber hinaus die Zitierfähigkeit (DOI) als auch die Auffindbarkeit der Forschungsergebnisse durch Suchmaschinen wie https://search.datacite.org/works?query=kwb. Zusätzlich lassen sich über vordefinierte Metadaten Publikationen zu EU Förderprojekten zuordnen, wodurch diese in OpenAire zum Projekt verlinkt werden (z.B. für das FAKIN Testprojekt AquaNES: https://explore.openaire.eu/search/project?projectId=corda__h2020::8c185265dbec7f46eddad1590900a00d).

Für die Projektadministration ergeben sich durch die Aufnahme des Forschungsdatenmanagements in das QMS folgende zusätzliche Verantwortlichkeiten für die Projektleiter:

- Ein **Datenmanagementplan** ist bei der Projektplanung zu berücksichtigen (verpflichtend bei EU H2020 Projekten) und bei Projektbeginn zu erstellen. Wenn nicht anders vorgegeben, ist hierzu das Online-Tool „DMPonline“ (https://dmponline.dcc.ac.uk/) zu nutzen
- Ein **eindeutiger Projekt-"Identifizierer"** ist zu Projektbeginn unter Beachtung der Namenskonventionen festzulegen und in den Ordner- und Dateinamen entsprechend den Vorgaben der Serverbereiche und Namenskonventionen zu verwenden. Dieser ist nicht zu verwechseln mit dem offiziellen Projektakronym.
- Es sind **regelmäßige "Datenreviews"** durchzuführen. Dabei ist die Einhaltung der oben aufgeführten "Best Practices" zu überprüfen. Dazu wurde in FAKIN ein Pfadanalysetool entwickelt (Sonnenberg & Rustler 2020), mit dem sich schnell Verletzungen der Nomenklatur-Vorgaben, fehlende Metadaten oder Datei-Duplikate identifizieren lassen. Darüber hinaus ermöglicht das Tool die
    - Visualisierung von Ordnergrößen und Anzahl von Dateien in sogenannten „Treemaps“ oder Dateibäumen.
    - Regelmäßige, automatisierte Erstellung von Dateilisten (z.B. über Nacht) zur Vereinfachung und Beschleunigung der Suche nach Dateien, da Textdateien anstatt Festplatten durchsucht werden. Darüber hinaus erlauben die Dateilisten auch einen "Blick zurück".

- **Veröffentlichungen**: in öffentlich geförderten Projekten ist am Ende der Laufzeit sicherzustellen, dass Projektberichte und ggf. auch Programmcode auf Zenodo (und letzterer darüber hinaus auch auf Github) veröffentlicht sind. Dieses Vorgehen soll dazu beitragen, dass das Ziel des KWB „mehr als 90 Prozent aller öffentlichen Projekt Berichte auf Zenodo zu veröffentlichen“ erreicht werden kann. Darüber hinaus ist die Ablage von Daten und Berichten in einem internen „Project closure document“ zu dokumentieren.

Arbeitspaket 3

Workshops (AP 3.1)

Auftaktworkshop

Im Rahmen einer KWB-internen Veranstaltung (sogenanntes „Brownbag“) am 7. Juli 2017 wurde das Thema Forschungsdatenmanagement sowie die Arbeiten und Ziele im Projekt FAKIN erstmals 16 Kollegen vorgestellt und diese auf den im September 2017 geplanten Auftaktworkshop eingestimmt (https://kwb-r.github.io/fakin/talk/2017-07-07_brownbag/).

Der interne Auftaktworkshop fand am 25. September 2017 im KWB statt. An dem Workshop nahmen 14 wissenschaftliche Mitarbeiter aus allen drei Fachabteilungen des KWB teil. Geleitet wurde der Workshop von den FAKIN-Mitarbeitern Michael Rustler und Hauke Sonnenberg. Der Workshop diente einerseits dazu, im Dialog mit den Mitarbeitern des KWB die verschiedenen Bereiche des Forschungsdatenmanagements (FDM) bezüglich ihrer Bearbeitung im Rahmen von FAKIN zu priorisieren. Andererseits sollten Vorschläge und Ideen zur Erreichung eines verbesserten FDM erarbeitet werden. Vorbereitend wurde eine Tabelle über mögliche Themenfelder des Datenmanagements aus der Literatur (Datenschule 2017; Sternkopf 2017) übernommen, an die Verhältnisse im KWB angepasst und in einen Online-Fragebogen (Anhang C) umgewandelt. Im ersten Teil des Workshops wurde der Fragebogen von den Teilnehmern online mit der Software LimeSurvey (LimeSurvey 2020) ausgefüllt und kurz ausgewertet. Folgende Schwerpunktthemen der Mitarbeiter wurden so identifiziert und priorisiert:

- Daten finden und bekommen,
- Daten analysieren, visualisieren (und kommunizieren),
- Daten säubern und verifizieren.

Im zweiten Teil des Workshops führten die Teilnehmer in Gruppenarbeit Brainstormings zu diesen drei Themenbereichen durch. Jede Gruppe erarbeitete eine Beschreibung des Ist-Zustandes, formulierte Ziele und entwickelte erste Ideen und Vorschläge zur Erreichung dieser Ziele, die auf Flipcharts dokumentiert wurden (Anhang D - Anhang E und https://kwb-r.github.io/fakin/talk/2017-09-25_workshop-kickoff/).

Best-Practices Workshop

Der Best-Practices Workshop wurde von Dezember 2017 bis Januar 2018 durch Interviews mit sechs Projektleitern des KWB vorbereitet. Ziel war es, einen genauen Überblick über den Ist- und den vom jeweiligen Projektleiter gewünschten Soll-Zustand zu Fragen des Datenmanagements zu erhalten. Gefragt wurde unter anderem:

- Mit welcher Art von Daten hast du (am meisten) zu tun (z.B. Zeitreihen vs. Einzelmessungen, zeitlich hoch aufgelöste vs. aggregierte Daten), Mit welchen Dateien (Dateitypen) hast du (am meisten) zu tun (z.B. MS Accessdatenbanken vs. Excel-Dateien vs. CSV-Dateien vs. Textdateien)?
- Von wem erhältst du Daten (z.B. extern: von Partnern, intern: von Messgeräten)
- Was sind deine häufigsten Probleme bei der Datenanalyse, wofür erhoffst du dir eine Lösung von FAKIN?
- Welche Verarbeitungsstufen von Daten (außer Rohdaten) kommen bei dir vor (z.B. rein formale Änderungen, Validieren, Bereinigen)?
- Findet sich der Datenverarbeitungsworkflow in der Datenstruktur wieder? Wenn ja, wie genau?

Die Interviews ermöglichten es, den Fokus auf die Best-Practices zu richten, die für die Wissenschaftler und Projektbearbeiter relevant sind. Zudem konnten so auch erfolgreiche Strategien von Kollegen in die Erarbeitung der Best-Practices einfließen.

Der KWB-interne Best-Practices Workshop fand am 29. Januar 2018 im KWB statt. An dem Workshop nahmen 13 Mitarbeiter aus allen drei Fachabteilungen des KWB teil. Geleitet wurde der Workshop von den FAKIN-Mitarbeitern Michael Rustler und Hauke Sonnenberg. Zu folgenden Themen wurden Best-Practices vorgestellt:

- **Ordnerstruktur:** Anwendung des EVA-Prinzips (Eingang, Verarbeitung, Ausgabe) durch Trennung von Rohdaten, Datenverarbeitung und Ergebnissen auf dem Server
- **Namenskonventionen** (z.B. Definition von zulässigen Dateinamen, Projektakronymen)
- **Versionierung**: mit / ohne technische Hilfsmittel (d.h. Versionsverwaltungssoftware wie z.B. Git oder Subversion)
- **Metadaten:** Definition von Mindestanforderungen (in einem Ordner sind alle Dateien / Unterordner durch deren Beschreibung in einer im YAML Format erstellten README Textdatei beschrieben).

Der Workshop diente auch dazu, Testprojekte zu identifizieren, an denen die praktische Anwendbarkeit der oben genannten Best-Practices überprüft werden sollte.

Es wurde sowohl das seit dem Jahr 2016 am KWB laufende EU-Projekt AquaNES (Projektseite), als auch das EU-Projekt Smart-Plant (Projektseite) ausgewählt. Letzteres stellt – obwohl ebenfalls bereits im Jahr 2016 gestartet, ein Beispiel für die Etablierung von Forschungsdatenmanagement in einem „neuen“ Projekt dar, da die aktive Projektbearbeitung am KWB erst zu Beginn der FAKIN Testprojektphase erfolgt ist. Die beiden Testprojekte unterscheiden sich darüber hinaus bezüglich ihrer FDM-Herausforderungen: Während im AquaNES-Projekt am KWB die Verarbeitung von großen Datenmengen von Pilotanlagen (eigener als auch von Projektpartnern) und Softwareentwicklung im Vordergrund steht, sind dies im Projekt Smart-Plant Life-Cycle Assessment Modellierungen und die Dokumentation von Abhängigkeiten zwischen LCA-Datenbank, Dateneingabe und Modellversionen. Durch diese thematische Unterschiedlichkeit ist gewährleistet, dass die Prüfung der Praxistauglichkeit der Best-Practices (vgl. AP 1.1) einen großen Bereich der am KWB in der inhaltlichen Projektbearbeitung auftretenden „Use-cases“ abdeckt.

Lessons-learnt Workshop

Der dritte interne “Lessons-learnt” Workshop fand am 11. März 2019 mit insgesamt 13 Kollegen aus allen drei Forschungsabteilungen, der Verwaltung und dem IT Dienstleister statt. Ziel war es, die Erfahrungen aus der über einjährigen Anwendung der „Best-practices“ in zwei Testprojekten vorzustellen und abzustimmen, welche Themen in das unternehmensweite Qualitätsmanagement zum Forschungsdatenmanagement aufgenommen werden sollten. Die Ergebnisse flossen direkt in die Bearbeitung der zugeordneten Task 2.2 ein (vgl. AP2).

Darüber hinaus wurden zwei im Rahmen von FAKIN entwickelte Werkzeuge zur Versbesserung des Forschungsdatenmanagements vorgestellt. Zum einen war dies das Pfadanalysetool, das zur Überprüfung der Einhaltung der aufgestellten „Best-practices“ (vgl. AP 1.1) im Rahmen der regelmäßig stattfindenden „Project-reviews“ eingesetzt werden kann. Zum anderen wurde ein Prototyp für einen institutionellen Wissensspeicher am KWB (vgl. AP3.3) vorgestellt und mittels einer LimeSurvey (2020) Online-Umfrage quantifiziert:

- Wie gut sich damit Fragen zu Personen, Publikationen, Werkzeugen oder Testprojekten beantworten lassen,
- Wie groß der Bedarf hierfür ist und
- Welche weiteren Fragen dieser idealerweise beantworten sollte.

Abschluss-Workshop

Der FAKIN-Abschlussworkshop wurde am KWB am 25. Juli 2019 mit 14 externen Teilnehmern (Anhang F) aus kleinen außeruniversitären Forschungsinstituten und Ingenieurbüros durchgeführt. Die Einladung erfolgte über die persönliche Ansprache bestehender Kontakte. Außerdem wurde die Deutsche Industrieforschungsgemeinschaft Konrad Zuse e.V. angeschrieben und deren Verteiler genutzt.

Im Workshop wurden der Projektablauf und die Ergebnisse für unser Institut in Form einer „Forschungsdatenreise“ dargestellt, und die zwei innerhalb von FAKIN entwickelten FDM-Werkzeuge (Pfadanalyse, Wissensspeicher), vorgestellt, deren Anwendung generalisierbar ist.

In einer Feedback-Runde wurden von den Teilnehmern Herausforderungen für das Forschungsdatenmanagement an kleinen Instituten genannt, beispielsweise:

Wenn keine explizite Förderung (Akquisition von Fördergeldern) existiert, dann hängt FDM an kleinen Instituten stark von dem Engagement von einzelnen Mitarbeitern ab (z.B. Entwicklung von automatisierten Routinen), die jedoch ggf. einer großen Fluktuation unterliegen

Es sollten nicht nur „verdichtete“ Daten langzeitgesichert werden, sondern auch die „Rohdaten“. Zudem wird die Wichtigkeit von Forschungsdatenmanagement meist erst erkannt, wenn etwas schiefläuft. Diese Gelegenheit sollte dann genutzt werden, um FDM voranzutreiben.

Zudem bekundeten zahlreiche Teilnehmer großes Interesse an dem im Rahmen des Projektes entwickelten generischen Pfadanalysetools, so dass geplant ist, dieses unter open-source Lizenz zu veröffentlichen und der FDM-Community zur Verfügung zu stellen.

Erfahrungsaustausch, Schulungen & Wissensskalierung (AP 3.3)

Die Kommunikation von Projektergebnissen erfolgte auf vielfältige Weise intern und extern.

KWB-interne Informationsveranstaltungen (sogenanntes „Brown-bag“)

Im Rahmen einer KWB internen Informationsveranstaltung (sogenanntes „Brownbag“) am 7. Juli 2017 wurde das Thema Forschungsdatenmanagement sowie die Arbeiten und Ziele im Projekt FAKIN erstmals 16 Kollegen vorgestellt und diese auf den im September 2017 geplanten Auftaktworkshop eingestimmt (https://kwb-r.github.io/fakin/talk/2017-07-07_brownbag/).

Erste Zwischenergebnisse aus den beiden Testprojekten wurden im Rahmen einer zweiten „Brown-Bag“ Veranstaltung (https://kwb-r.github.io/fakin/talk/2018-09-21_brownbag/) am 21. September 2018 durch die FAKIN Mitarbeiter Michael Rustler und Hauke Sonnenberg 10 Kollegen aus allen Abteilungen vorgestellt. Es konnte beispielhaft für das AquaNES Projekt gezeigt werden, wie sich eine verbessere Ordnerstruktur auf das Finden von Dateien auswirkt: in der neuen Struktur ließen sich auf direkterem Weg, in kürzerer Zeit und fast in allen Fällen die gewünschten Daten finden (siehe auch Anhang B).

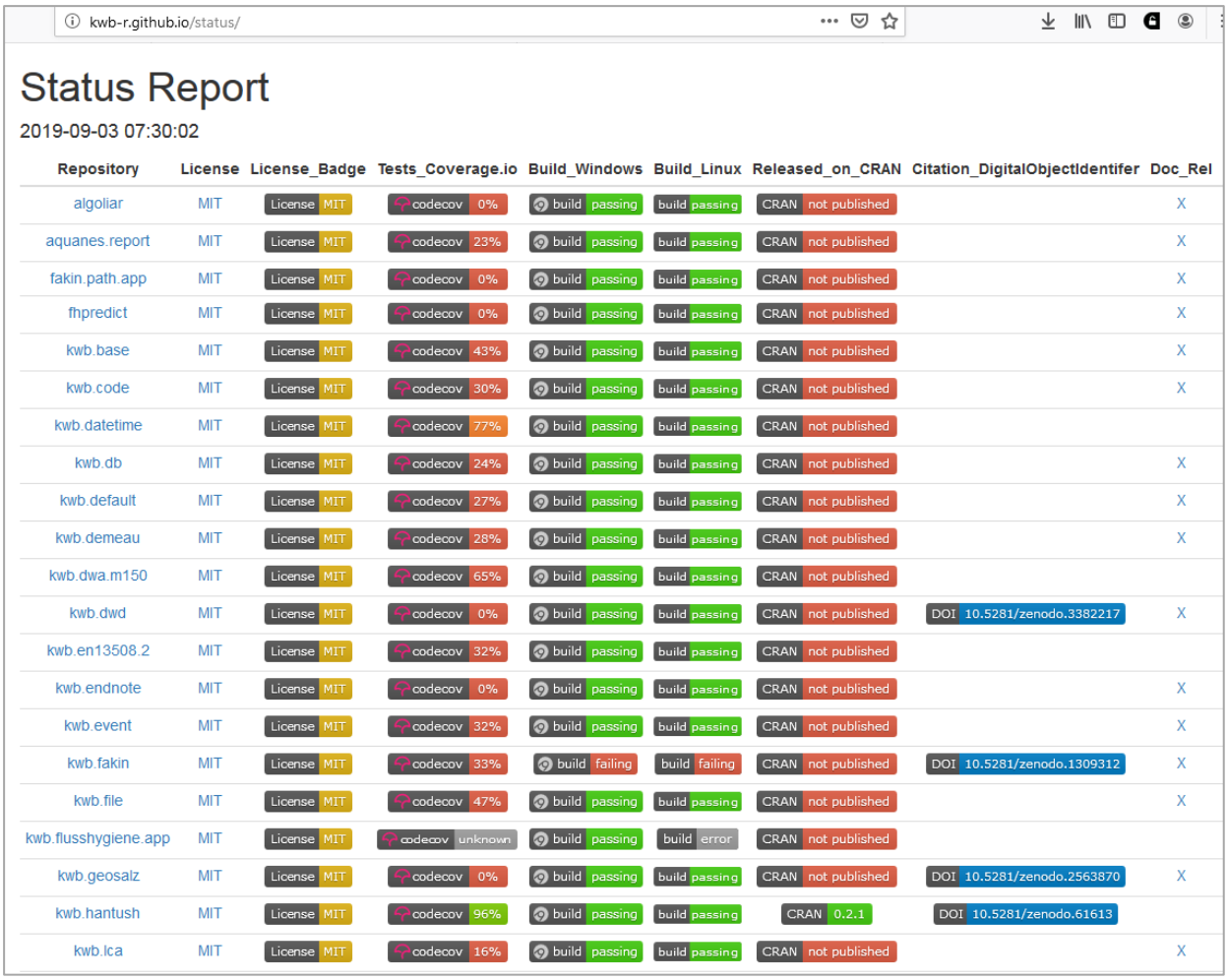
Veröffentlichung von Code auf GitHub und Zenodo

Insgesamt wurden auf der kollaborativen Code Plattform GitHub im FAKIN-Projektzeitraum 2017-2019 über 34 R-Pakete (Anhang G) unter der offenen MIT-Lizenz veröffentlicht. Darüber hinaus wurde eine Integration zwischen GitHub und dem zur Langzeitarchivierung geeigneten Repositorium Zenodo geschaffen (https://guides.github.com/activities/citable-code/), der für jeden „Release“ auf GitHub, diesen automatisch auf Zenodo kopiert und einen Digital Object Identifier (DOI) anlegt. Insgesamt wurden im FAKIN Projektzeitraum 2017-2019 acht R-Pakete auf Zenodo zitierbar gemacht und langzeitarchiviert (https://zenodo.org/communities/kwb, Anhang H).

Auf GitHub wurden folgende Funktionen und Dienste genutzt:

- Kontinuierliche Integration: testet ob sich die Software auf verschiedenen Betriebssystemen problemlos installieren lässt. Hierzu wurden die für öffentliche open-source Projekte kostenlosen Angebote der Cloud-Dienstleister Travis-CI (2020) und Appveyor (2020) genutzt. Funktionstests: prüft ob die Software den Erwartungen nach funktioniert. Hierzu wird das für öffentliche open-source Projekte kostenlose Angebot des Cloud-Anbieters Codecov (2020) genutzt (z.B. https://codecov.io/gh/KWB-R/kwb.qmra)
- (Software-)Projektmanagement: d.h. Erstellung von Issues, Milestones und deren Zusammenfassung in einem Projekt (https://github.com/KWB-R/kwb.qmra/projects/1).
- Bereitstellung/Veröffentlichung von interaktiven Anwendungen: mit Hilfe des kostenlosen Dienstes MyBinder (2020), siehe beispielsweise: https://mybinder.org/v2/gh/kwb-r/apps/aquanes.report?urlpath=shiny/haridwar/
- Automatisierte Dokumentation (siehe z.B. https://kwb-r.github.io/aquanes.report/): durch Kombination von Travis-CI (2020) zur Dokumentationserstellung, GitHub Pages (2020) zum Webhosting und der für open-source Projekte kostenlosen Suchfunktion „DocSearch“ der Firma Algolia (https://community.algolia.com/docsearch/).

Um eine strukturierte Übersicht über alle veröffentlichten R Pakete des KWB zu erhalten, wurde das R-Paket „kwb.pkgstatus“ (Rustler 2020a) erstellt, mit dessen Hilfe einmal täglich eine Tabelle mit verschiedenen Statusinformationen aktualisiert wird (Abbildung 5).



kwb-r.github.io/status/

Status Report

2019-09-03 07:30:02

Repository	License	License_Badge	Tests_Coverage.io	Build_Windows	Build_Linux	Released_on_CRAN	Citation_DigitalObjectIdentifer	Doc_Rel
algoliar	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published		X
aquanes.report	MIT	License MIT	codecov 23%	build passing	build passing	CRAN not published		X
fakin.path.app	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published		X
fhpredict	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published		X
kwb.base	MIT	License MIT	codecov 43%	build passing	build passing	CRAN not published		X
kwb.code	MIT	License MIT	codecov 30%	build passing	build passing	CRAN not published		X
kwb.datetime	MIT	License MIT	codecov 77%	build passing	build passing	CRAN not published		
kwb.db	MIT	License MIT	codecov 24%	build passing	build passing	CRAN not published		X
kwb.default	MIT	License MIT	codecov 27%	build passing	build passing	CRAN not published		X
kwb.demeau	MIT	License MIT	codecov 28%	build passing	build passing	CRAN not published		X
kwb.dwa.m150	MIT	License MIT	codecov 65%	build passing	build passing	CRAN not published		
kwb.dwd	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published	DOI 10.5281/zenodo.3382217	X
kwb.en13508.2	MIT	License MIT	codecov 32%	build passing	build passing	CRAN not published		
kwb.endnote	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published		X
kwb.event	MIT	License MIT	codecov 32%	build passing	build passing	CRAN not published		X
kwb.fakin	MIT	License MIT	codecov 33%	build failing	build failing	CRAN not published	DOI 10.5281/zenodo.1309312	X
kwb.file	MIT	License MIT	codecov 47%	build passing	build passing	CRAN not published		X
kwb.flusshygiene.app	MIT	License MIT	codecov unknown	build passing	build error	CRAN not published		
kwb.geosalz	MIT	License MIT	codecov 0%	build passing	build passing	CRAN not published	DOI 10.5281/zenodo.2563870	X
kwb.hantush	MIT	License MIT	codecov 96%	build passing	build passing	CRAN 0.2.1	DOI 10.5281/zenodo.61613	
kwb.lca	MIT	License MIT	codecov 16%	build passing	build passing	CRAN not published		X

Abbildung 5 Täglich aktualisierter Statusbericht der auf GitHub veröffentlichten R Pakete (https://kwb-r.github.io/status)

Durch die Nutzung der GitHub Plattform und weiterer Dienste (Travis-CI, R Paket „pkgdown“) können für die dort veröffentlichten R-Pakete automatisiert Dokumentationswebseiten erstellt werden. Dieser Schritt ermöglicht es, Online-Tutorials wie z.B. mit dem Thema „Wie installiere ich KWB Pakete von GitHub“ (https://kwb-r.github.io/kwb-r/kwb.pkgbuild/articles/install.html) bereitzustellen. Zusätzlich lassen sich die Dokumentationen dank einer Integration mit dem für Codedokumentationen kostenfreien Dienst „DocSearch“ der Firma Algolia (https://community.algolia.com/docsearch) schnell durchsuchen, wie zum Beispiel: https://kwb-r.github.io/kwb.pkgbuild/reference/use_installation.html?q=install

FAKIN Wissensspeicher (Knowledge Repo)

Die oben genannten Veröffentlichungen von vorher meist nur intern genutztem Programmcode auf GitHub stellten einen vorbereitenden Schritt zur Erstellung eines unternehmensweiten „Knowledge Repos“ (Wissensspeichers) dar. Die im Projektantrag genannte ursprüngliche Idee, hierfür die von AirBnB als open-source veröffentlichte Software „Knowledge Repo“ (AirBnB 2018) zu nutzen, wurde verworfen, da dessen Fokus auf den Wissensaustausch von Programmierern oder Data-Scientists beschränkt ist (AirbnbEng 2016; Bion et al. 2018).

Daher wurde eine Eigenentwicklung präferiert, die im Rahmen von FAKIN als Proof-of-Concept umgesetzt wurde. Die Idee war es, die am KWB eingesetzte open-source Programmiersprache R (R Core Team 2020) mit weiterer open-source Software wie dem R Paket „blogdown“ (Xie 2019), dem Webseitengenerator Hugo (2019) und weiteren kostenfreien Diensten und Software zu kombinieren, um damit eine statische Website zu erstellen in der das Wissen aus verschiedenen Datenquellen integriert werden kann (Rustler et al. 2019b).

Bis Anfang März 2019 wurde ein Wissensspeicher für das FAKIN-Projekt als Prototyp für einen institutsweiten Wissensspeicher als statische Website erstellt (Rustler 2019a). Dieser enthält Informationen aus den folgenden Datenquellen (Abbildung 6):

- Programmcode (GitHub),
- Publikationen (Zenodo) und
- Institutsinternen Informationen zu Personen, Testprojekten, Tools und Vorträgen (erfasst in standardisierten CSV Dateien)

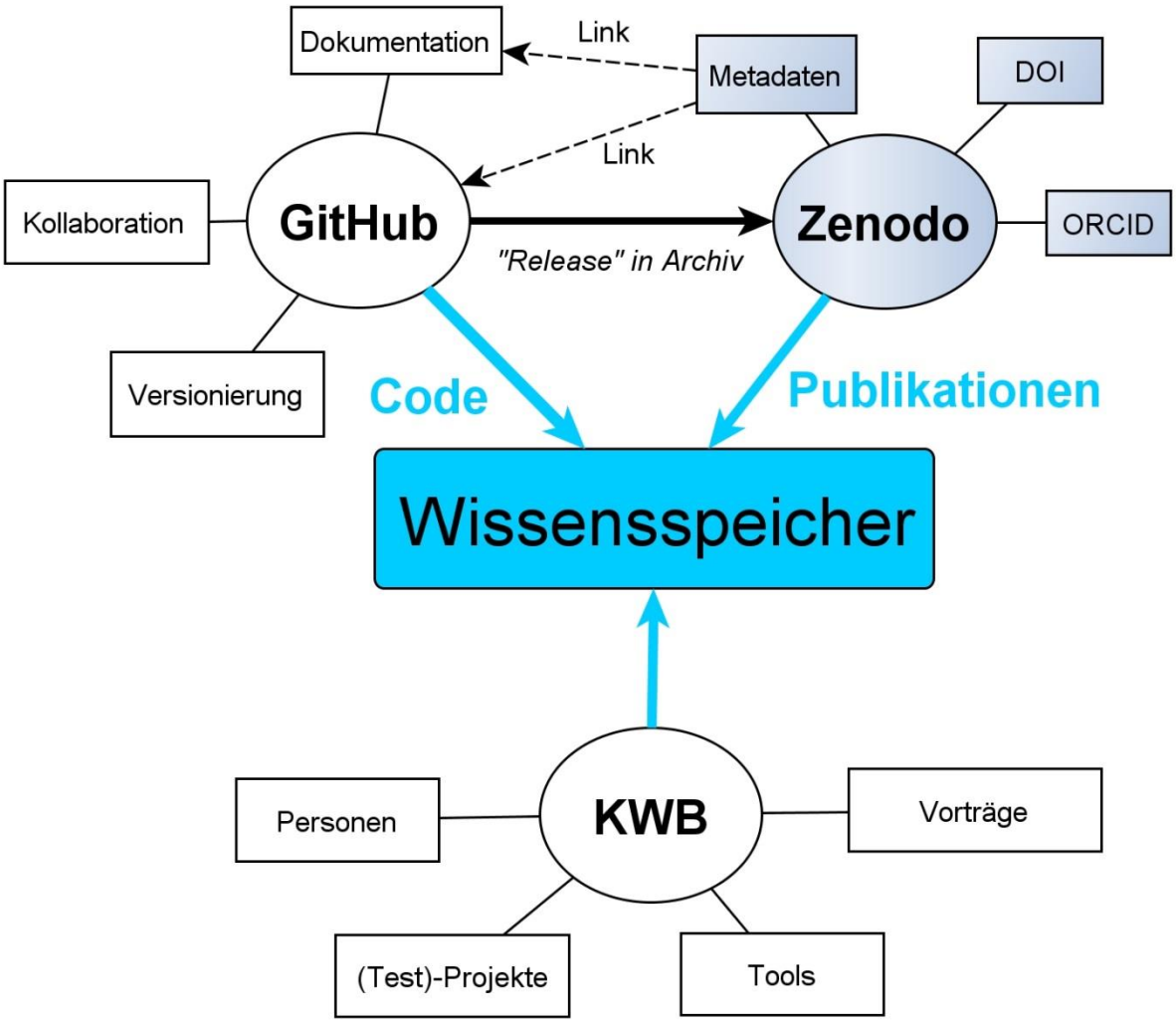


Abbildung 6 Datenquellen für den FAKIN Wissensspeicher Prototyp: Github (Programmcode), Zenodo (Publikationen) und institutsinternen Wissen (Personen, Projekte, Tools, Vorträge).

Abbildung 7 zeigt die Startseite des FAKIN Wissensspeichers. Bei Auswahl des Menüpunktes „Tools“ werden z.B. alle Werkzeuge, die in FAKIN getestet wurden, thematisch filterbar angezeigt, wie am Beispiel der Umfragetools gezeigt (Abbildung 8). Interessiert man sich im nächsten Schritt für ein bestimmtes Umfragetool, z.B. LimeSurvey (Abbildung 9), so sind in der Detailansicht neben einer Kurzbeschreibung nicht nur die Personen verlinkt, die sich damit auskennen, sondern auch z.B., bei welchen FAKIN-Workshops es verwendet wurde (https://kwb-r.github.io/fakin/tool/limesurvey/).

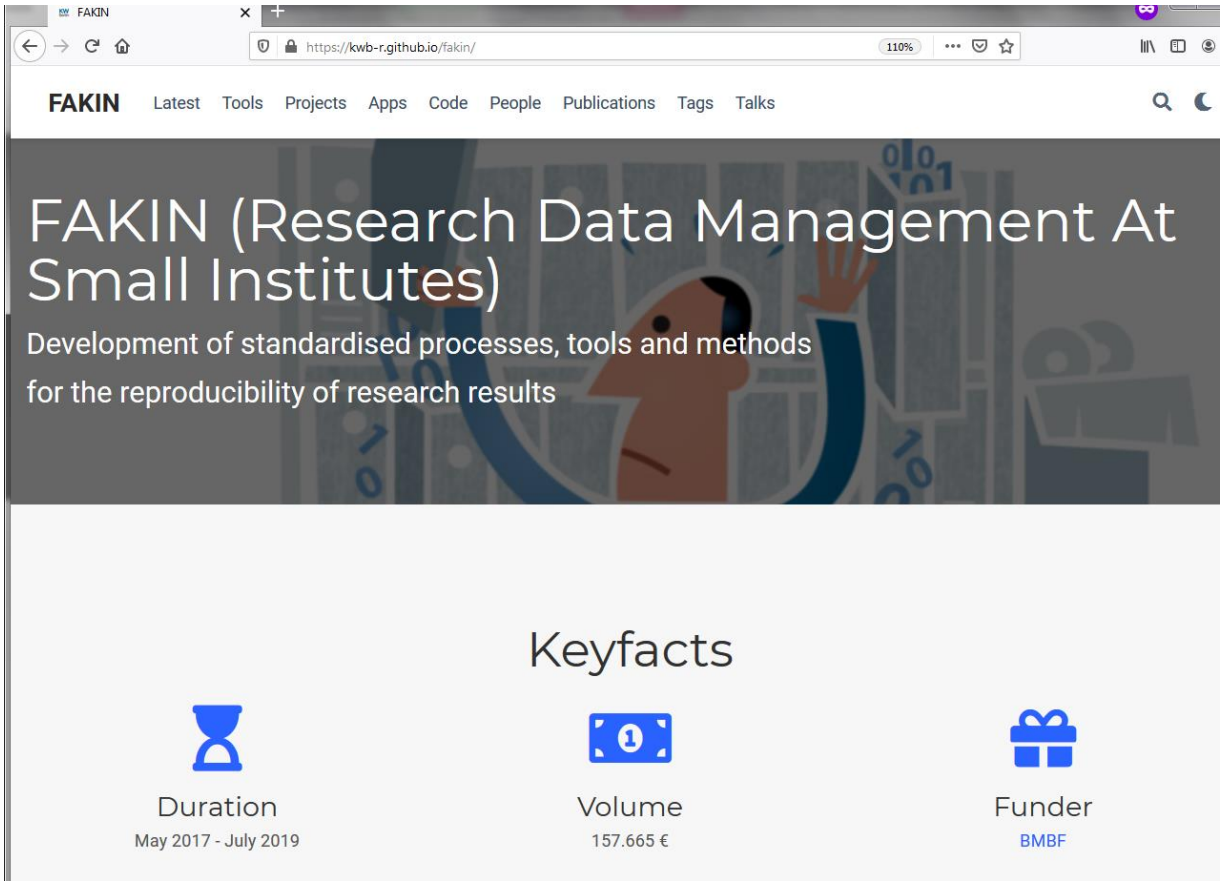


Abbildung 7 **Screenshot der Startseite des FAKIN Wissenspeicher Prototyps (https://kwb-r.github.io/fakin)**

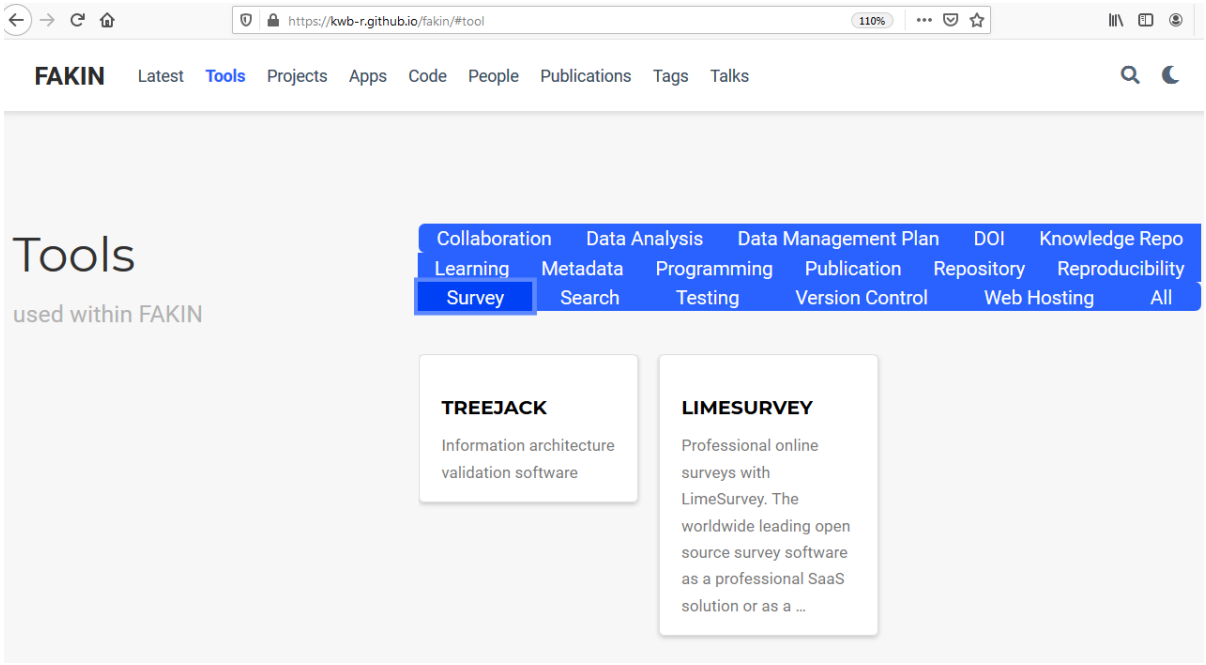


Abbildung 8 **Screenshot des FAKIN Wissenspeichers mit den im Projekt getesteten Tools, die thematisch gefiltert werden können (https://kwb-r.github.io/fakin/#tool). Hier dargestellt sind die beiden Umfragetools (Treejack, Limesurvey)**

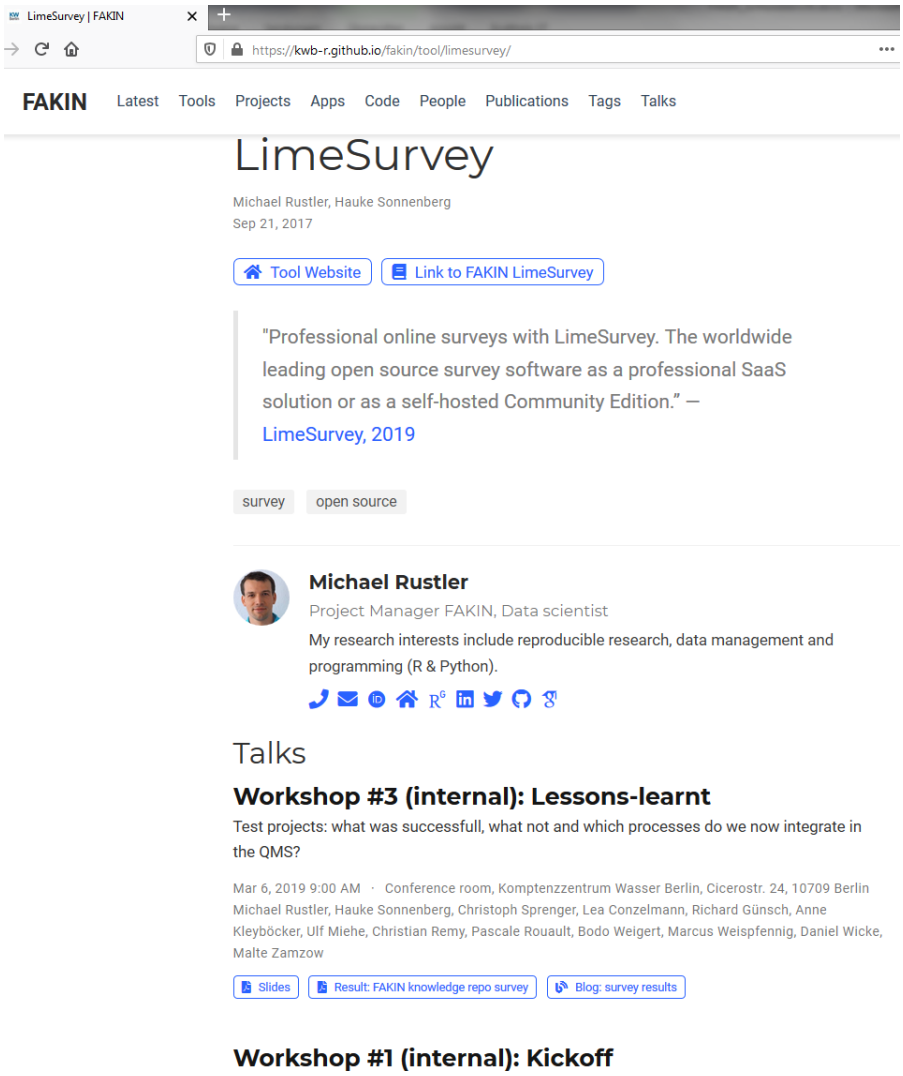


Abbildung 9 Detailansicht für Tool „LimeSurvey“ (https://kwb-r.github.io/fakin/tool/limesurvey) mit Kurzbeschreibung, Personen die sich damit auskennen und bei welchen Workshops es zum Einsatz kam

Der Wissensspeicherprototyp wurde im Rahmen des Lessons-learnt Workshops am 11. März 2019 von 17 Kolleg(inn)en getestet. Innerhalb von 30 Minuten waren:

- sieben verschiedene Fragen mit Hilfe des Wissensspeichers zu beantworten (***76% richtige Antworten***),
- die Nutzbarkeit im Schulnotensystem zu bewerten (***Notendurchschnitt 2,5***) und abschließend die
- Wichtigkeit einer Skalierung des FAKIN-Wissensspeichers auf das gesamte Institut einzuschätzen (85 Prozent der Befragten mindestens „Wichtig“ – 77 Prozent – oder sogar „Sehr Wichtig“ – 8 Prozent).

Diese Befragung wurde mit dem open-source Tool LimeSurvey durchgeführt und die Ergebnisse in einem Blogpost im FAKIN-Wissensspeicher detailliert zusammengefasst und veröffentlicht (Rustler 2019b).

II.2 Die wichtigsten Positionen des zahlenmäßigen Nachweises

Die Zuwendungen wurden am KWB insbesondere für Personalmittel verwendet:

- 15% für die Arbeiten im Arbeitspaket 1
- 42% für die Arbeiten im Arbeitspaket 2
- 43% für die Arbeiten im Arbeitspaket 3

Sachkosten fielen nur für die Durchführung des Abschlussworkshops an.

Die folgende Tabelle 2 zeigt eine Zusammenfassung des zahlenmäßigen Nachweises für das KWB.

Tabelle 2: Zahlenmäßiger Nachweis KWB

Position	Gesamtvorkalkulation [€]	Gesamtnachkalkulation [€]
0813 Material	1.500,-- €	327,35 €
0837 Personalkosten	153.166,-- €	157.395,81 €
0838 Reisekosten	3.000,-- €	0 €
0881 gesamte Selbstkosten des Vorhabens	157.666,-- €	157.723,16 €
0882 Eigenmittel des Antragstellers	31.533,20 €	31.590,36 €
0884 Zuwendung (Förderquote 80%)	126.132,80 €	126.132,80 €

II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die finanziellen Mittel und die Arbeiten wurden entsprechend den im Antrag dargelegten Arbeitspaketen geleistet und an die im Verlauf des Projektes erhaltenen Ergebnisse ausgerichtet.

II.4 Darstellung des voraussichtlichen Nutzens

Wissenschaftlich-technische und wirtschaftliche Erfolge

Das Projekt FAKIN hat wesentlich zur institutionellen Verankerung des Forschungsdatenmanagements am KWB beigetragen, insbesondere durch die Einbindung der Mitarbeiter in die Ist-Soll-Analysen und die Übernahme von Best Practices und Workflows in das unternehmensweite Qualitätsmanagementsystem. Als besonderer wissenschaftlich-technischer Erfolg ist die Entwicklung von zwei komplett auf open-source Software basierten FDM – Werkzeugen zu sehen:

Zum einen wurde ein generisches Pfadanalysetool entwickelt, mit dem sich kontinuierlich die Einhaltung der aufgestellten Best-Practices zu Namenskonventionen sowohl quantitativ als auch visuell mittels Ampelsystem überprüfen lässt. Aufgrund des positiven Feedbacks und Interesses wurde dieses Pfadanalysetool unter der permissiven MIT-Lizenz auf Zenodo (Sonnenberg & Rustler 2020) veröffentlicht, so dass eine Nutzung, Anpassung oder Weiterentwicklung durch die FDM-Community möglich ist.

Zum anderen wurde ein Projektwissensspeicher aufgebaut, der Informationen vernetzt darstellt (Personen, Werkzeuge, Vorträge, Publikationen) und durch Nutzung einer fortschrittlichen SaaS (Software-as-a-Service) Suchmaschine der Firma Algolia schnell und effizient durchsuchbar ist. Diese ermöglicht eine Volltextsuche, wobei bereits mit der ersten Buchstaben-Eingabe – ähnlich wie bei Google – passende Suchergebnisse in Echtzeit geliefert werden. Darüber hinaus ist das Einpflegen von Wissen voll- (z.B. Abgreifen von Informationen aus öffentlichen Repositorien: Zenodo, GitHub) oder teilautomatisiert möglich (Einlesen über vordefinierte CSV Textdateien: z.B. Informationen über Mitarbeiter). Damit stellt der FAKIN Wissensspeicher einen Prototyp für einen unternehmensweiten Wissensspeicher dar, den sich viele Kollegen wünschen (Feedback Umfrage).

Wissenschaftlich-technische und wirtschaftliche Anschlussfähigkeit

Durch die Integration der in FAKIN erarbeiteten Prozesse, Best-Practices und Werkzeuge ins Qualitätsmanagementsystem des KWB wird sich mittelfristig das Forschungsdatenmanagement (FDM) am KWB verbessern, da nun erstmals für unser Institut verbindliche Arbeitsanweisungen vorliegen. Bereits in den drei internen Workshops konnten die Forscher(inn)en für FDM sensibilisiert werden, so dass diese Aspekte nun verstärkt auch in die Projektentwicklung eingebracht werden (z.B. aktuell in verschiedenen laufenden Anträgen: Nutzung der öffentlichen Codeplattform GitHub für im Projekt zu entwickelnden Programmcode mit Partnern).

Beteiligung des KWB an EU Ausschreibung zum Thema Digitalisierung

Das KWB konnte als Koordinator eines europäischen Konsortiums im Rahmen von EU-Horizon2020 Fördergelder in Höhe 4,99 Millionen Euro einwerben. In der Antragsphase des Projektes „Digital solutions for water: linking the physical and digital world for water solutions“ wurden Elemente aus FAKIN zu den Themen „Data Management Plan“ und „Metadaten Standards“ eingebracht. In FAKIN konnte ein Softwarewerkzeug entwickelt werden, welche die Verhandlungen mit den Projektpartnern über Budgetfragen vereinfachte. Dieses Werkzeug wird jetzt auch bei anderen Projektentwicklungen genutzt.

Im Laufe des Projektes wird außerdem die Erstellung und fortlaufende Überprüfung und Anpassung von Datenmanagementplänen durch die Arbeiten und Ergebnisse in FAKIN unterstützt.

Veröffentlichungen

Durch die systematische Veröffentlichung vieler zuvor nur intern am KWB verfügbarer Tools auf GitHub (https://kwb-r.github.io/status,https://github.com/kwb-r) und Zenodo (https://zenodo.org/communities/kwb) hat sich die Sichtbarkeit des KWB und dessen Engagement, die in der Forschung verwendeten Algorithmen überprüfbar und nachnutzbar (permissive MIT Lizenz) zu machen, verbessert. Damit können sich mittel- bis langfristig Kooperationen (Einbringen dieser Kompetenzen in Forschungsprojekte) oder Aufträge (z.B. Anpassung von Tools wie dem Pfadanalysetool für andere Forschungsinstitute/Partner).

Kompetenzzentwicklung Datenverarbeitung

Die bei der Entwicklung des Wissensspeicher-Prototyps gewonnenen Kompetenzen werden derzeit (<1 Jahr nach Projektende) zur Erstellung einer Instituts-Publikationswebseite eingebracht. Durch die verbesserten Datenverarbeitungskompetenzen ist es möglich, diese ohne externe Ausschreibung nur durch eigenes Personal am KWB zu realisieren. Ein entsprechender Prototyp für eine institutionelle Publikationswebsite wurde bereits im Rahmen von FAKIN umgesetzt und ist unter https://kwb-r.gitlab.io/pubs einsehbar.

II.5 Während der Durchführung des Vorhabens bekannt gewordene Fortschritte auf dem Gebiet des Vorhabens bei anderen Stellen

Während der Laufzeit des Projektes wurden keine Fortschritte auf den vom KWB erarbeiteten Gebieten durch andere Stellen erzielt, die den bisher erzielten Ergebnissen widersprechen bzw. weitergehende Aussagen zulassen.

Interessant ist ein dreijähriges FDM-Beratungsprojekt, das gemeinsam von der Sächsische Landesbibliothek - Staats- und Universitätsbibliothek Dresden (SLUB) und der TU Dresden durchgeführt wird und auf dem Workshop „Forschungsdatenmanagement praktikabel gestalten“ des Herder Instituts am 20.-21. August 2019 in Marburg von Andreas von der Dunk (SLUB) vorgestellt wurde. Dieses bietet FDM-Beratung für Projektvorhaben der Dresdner Hochschulen an, wobei bisher (Stand: August 2019) bereits mehr als 21 Beratungsgespräche geführt wurden. Ziel ist es, basierend auf den Beratungsergebnissen ein FDM-Tool zu entwickeln. In einem persönlichen Gespräch mit Andreas von der Dunk (SLUB), das im Anschluss an den Workshop stattfand, wurde vereinbart, ihn bei Veröffentlichung des Pfadanalysetools zu informieren, damit er testen kann, wie sinnvoll dessen Einsatz in der FDM-Beratung ist.

II.6 Erfolgte oder geplante Veröffentlichungen der Ergebnisse

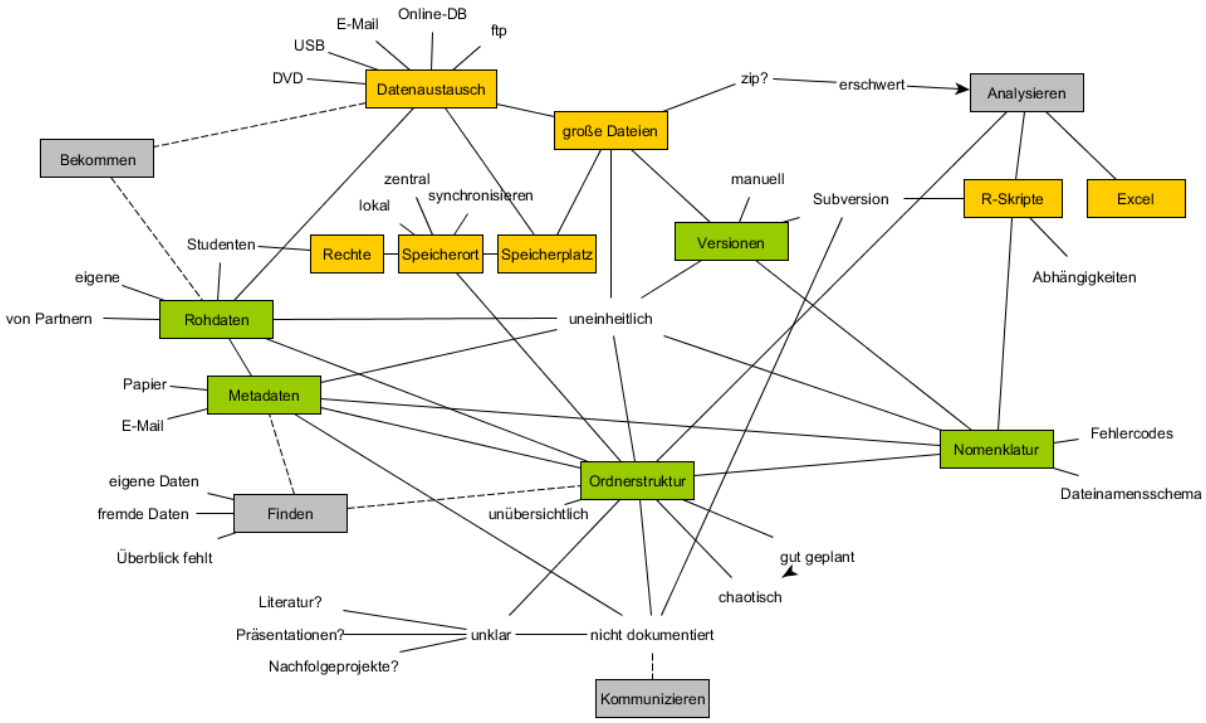
Die folgende Tabelle 3 zeigt die im Rahmen von FAKIN erfolgten Veröffentlichungen.

Tabelle 3: Veröffentlichungen nach Kategorie (Bericht, Blogpost, Code, Website)

Bericht	Rustler, M., Sonnenberg, H. & Sprenger, C. (2019) Best Practices in Research Data Management. Kompetenzzentrum Wasser Berlin gGmbH https://kwb-r.github.io/fakin.doc/ (abgerufen am: 2019-12-19)
Blogpost	Rustler, M. (2019) Lessons-learnt: Auswertung der FAKIN Wissensspeicher Umfrage Blogpost, https://kwb-r.github.io/fakin/post/workshop-lessons-learnt/ (abgerufen am: 2019-12-19)
Code	Boettiger, C., Salmon, M., Sonnenberg, H., Ross, N., Leinweber, K., Smith, A., Meyer, S., Rustler, M., Woo, K. & Krystalli, A. (2019) codemetar (v.0.1.7): Generate CodeMeta Metadata for R Packages, v.0.1.7, Zenodo, doi:10.5281/zenodo.2598516
Code	Rustler, M. (2019) fakin (v1.0): Source Code for Knowledge Repo of FAKIN Project, v1.0, doi:10.5281/zenodo.3603922
Code	Rustler, M. (2020) kwb.endnote (v0.1.0): R Package for Analysing KWB Endnote Library (Exported as .XML), v0.1.0, Zenodo, doi:10.5281/zenodo.3604072
Code	Rustler, M. (2020) kwb.pkgstatus (v0.1.0): R Package for Checking KWB Package Status v0.1.0, Zenodo, doi:10.5281/zenodo.3604014
Code	Rustler, M. (2020) kwb.pubs (v0.1.0): R Package for Generating Publications Website Based on Hugo Academic Theme v0.1.0, Zenodo, doi:10.5281/zenodo.3604178
Code	Rustler, M. (2020) pkgmeta (v0.1.0): R Package for Meta-Analysis of KWB-R Packages on GitHub, v0.1.0, Zenodo, doi:10.5281/zenodo.3604093
Code	Rustler, M. & Sonnenberg, H. (2019) kwb.pkgbuild (v0.1.2): R Package for Standardised Development at KWB, v0.1.2, Zenodo, doi:10.5281/zenodo.3430683
Code	Rustler, M. & Sonnenberg, H. (2020) kwb.umberto (v0.1.0): R Package for Supporting LCA with UMBERTO Software at KWB, v0.1.0, Zenodo, doi:10.5281/zenodo.3604007
Code	Sonnenberg, H. (2020) kwb.pathdict (v0.1.1): R-Package to Work with Path Dictionaries, v0.1.1, Zenodo, doi:10.5281/zenodo.3604196
Code	Sonnenberg, H. (2020) kwb.readxl (v0.1.0): R Package to Read Data from Excel Files v0.1.0, Zenodo, doi:10.5281/zenodo.3604187
Code	Sonnenberg, H. (2020) kwb.file (v0.3.0): R Package With Functions Related to File and Path Operations, Zenodo, v0.3.0, doi:10.5281/zenodo.3603493
Code	Sonnenberg, H. & Rustler, M. (2020) kwb.lca (v0.1.0): R Package for Life Cycle Assessment (LCA) Projects, v0.1.0, Zenodo, doi:10.5281/zenodo.3604184
Code	Sonnenberg, H. & Rustler, M. (2020) fakin.path.app (v0.3.0): R Shiny Application for Path Analysis, v.0.3.0, Zenodo, doi:10.5281/zenodo.3603503
Website	Rustler, M. (2019) FAKIN Wissensspeicher – ein proof-of-concept Beispiel für einen institutionellen Wissensspeicher. https://kwb-r.github.io/fakin (abgerufen am: 2019-12-19)

III. ANHANG

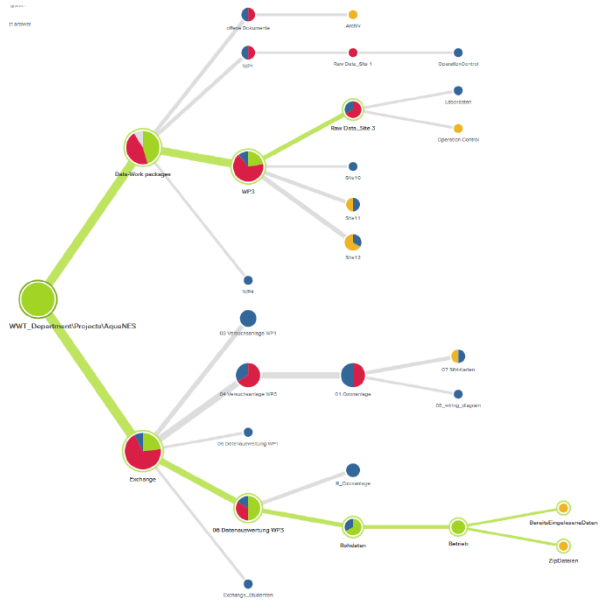
Anhang A **Mind Map Mitarbeiterinterviews**



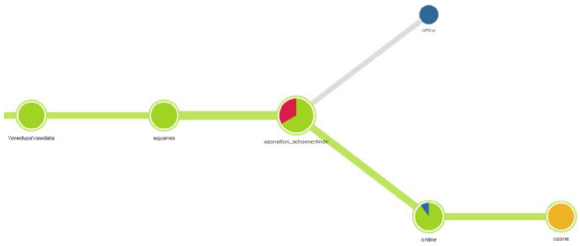
Anhang B: **Testprojekt: Aquanes: Optimierung der Ordnerstruktur „Dateien finden“ (Ergebnisse der Befragung von 9 KWB Wissenschaftlern mit Hilfe des Tools Treejack, OptimalWorkshop (2019))**

Wo finde ich die Rohdaten des Betriebes der Ozonanlage in Schönerlinde, die im Projekt AquaNes erhoben wurden?

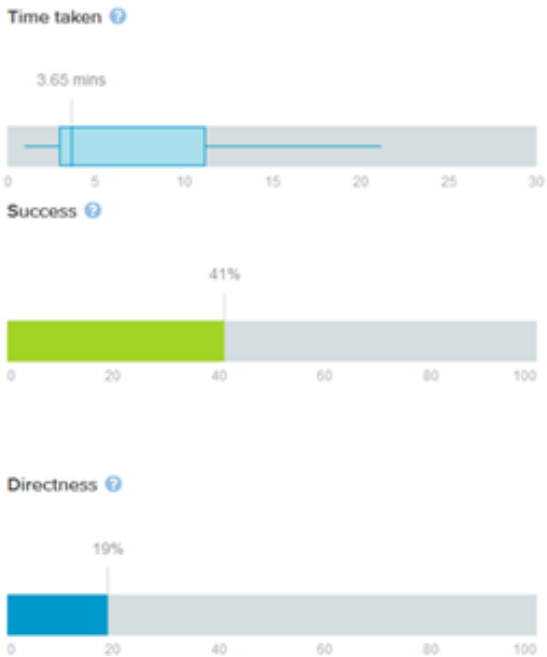
Bisherige Ordnerstruktur



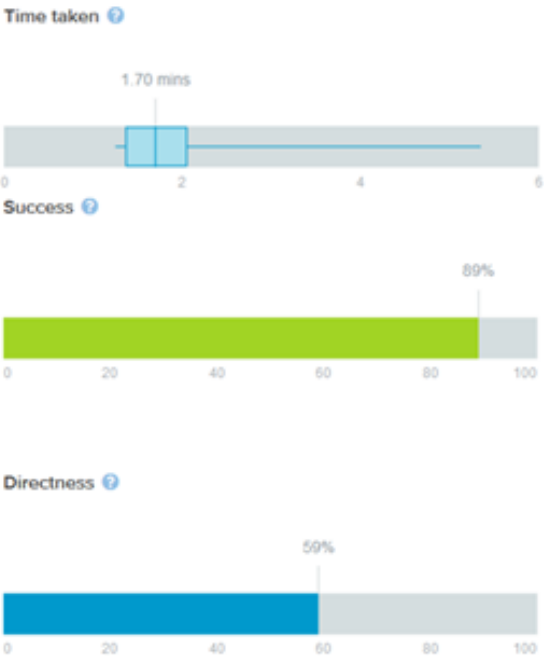
Neue Ordnerstruktur



Bisherige Ordnerstruktur n = 9



Neue Ordnerstruktur n = 9



Anhang C: **Auftaktworkshop Limesurvey Fragebogen**

Teil A: Individuelle Ebene: 1 - Daten finden

A1. IST-ZUSTAND: Welche Aussage beschreibt am besten deine Vorgehensweise, Daten auf dem KWB-Server (den KWB-Servern) zu finden?

1) Ich finde mich in den Projektordnern nicht zurecht. Daher frage ich eine/n Mitarbeiter/in, von der/dem ich glaube, dass sie/er die Daten erhoben bzw. abgelegt hat. ☐

2) Ich finde mich in den Projektordnern nicht zurecht. Aber mit Hilfe der Windows -Suche bin ich meistens in der Lage, die Dateien zu finden. ☐

3) Ich suche in den verschiedenen Unterordnern der Projektbaumstruktur nach der Zieldatei, die ich jedoch oftmals erst nach mehreren Anläufen und ggf. unter Zuhilfenahme weiterer Tools (z. B. Windows-Suche) finde. ☐

4) Die Daten liegen gut strukturiert auf dem Server. Eine einheitliche Benennung von Ordnern und Dateien sorgt dafür, dass ich sie meist intuitiv und ohne Verwendung von Suchmaschinen (z. B. Windows-Suche) auffinde. ☐

A2. SOLL-ZUSTAND: Daten finden

1) Ich finde mich in den Projektordnern nicht zurecht. Daher frage ich eine/n Mitarbeiter/in, von der/dem ich glaube, dass sie/er die Daten erhoben bzw. abgelegt hat. ☐

2) Ich finde mich in den Projektordnern nicht zurecht. Aber mit Hilfe der Windows -Suche bin ich meistens in der Lage, die Dateien zu finden. ☐

3) Ich suche in den verschiedenen Unterordnern der Projektbaumstruktur nach der Zieldatei, die ich jedoch oftmals erst nach mehreren Anläufen und ggf. unter Zuhilfenahme weiterer Tools (z. B. Windows-Suche) finde. ☐

4) Die Daten liegen gut strukturiert auf dem Server. Eine einheitliche Benennung von Ordnern und Dateien sorgt dafür, dass ich sie meist intuitiv und ohne Verwendung von Suchmaschinen (z. B. Windows-Suche) auffinde. ☐

Teil B: Individuelle Ebene: 2 - Daten bekommen

B1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze zum Umgang mit Daten, die du von externen Partnern (z. B. BWB) erhalten hast?

1) Wenn ich Daten per E-Mail erhalte, belasse ich sie im Posteingang, weil ich sie da immer finde. Zusätzlich lege ich sie lokal auf meinem Computer ab. ☐

2) Ich lege die Daten in einem Unterordner "Rohdaten_HIER_DRIN_NICHT_ARBEITEN" im entsprechenden Projektordner "Data-Work packages" ab. ☐

3) Ich lege die Daten in einem Unterordner "Rohdaten" im entsprechenden Projektordner "Data-Work packages" ab. Darüber hinaus stelle ich sicher, dass die Daten nicht verändert werden können (z. B. Schreibschutz). ☐

4) Ich lege die Daten in einem Unterordner "Rohdaten" im entsprechenden Projektordner "Data-Work packages" ab. Darüber hinaus stelle ich sicher, dass die Daten nicht verändert werden können (z. B. Schreibschutz). Zusätzlich lege ich eine Datei mit Metadaten ab. Diese gibt Auskunft darüber, woher die Daten stammen, ob und welche Nutzungseinschränkungen es gibt, wie die Daten erhoben wurden (z. B. Analysemethode bei Labormessungen), usw. ☐

B2. SOLL-Zustand: Daten bekommen

1) Wenn ich Daten per E-Mail erhalte, belasse ich sie im Posteingang, weil ich sie da immer finde. Zusätzlich lege ich sie lokal auf meinem Computer ab. ☐

2) Ich lege die Daten in einem Unterordner "Rohdaten_HIER_DRIN_NICHT_ARBEITEN" im entsprechenden Projektordner "Data-Work packages" ab. ☐

3) Ich lege die Daten in einem Unterordner "Rohdaten" im entsprechenden Projektordner "Data-Work packages" ab. Darüber hinaus stelle ich sicher, dass die Daten nicht verändert werden können (z. B. Schreibschutz). ☐

4) Ich lege die Daten in einem Unterordner "Rohdaten" im entsprechenden Projektordner "Data-Work packages" ab. Darüber hinaus stelle ich sicher, dass die Daten nicht verändert werden können (z. B. Schreibschutz). Zusätzlich lege ich eine Datei mit Metadaten ab. Diese gibt Auskunft darüber, woher die Daten stammen, ob und welche Nutzungseinschränkungen es gibt, wie die Daten erhoben wurden (z. B. Analysemethode bei Labormessungen), usw. ☐

Teil C: Individuelle Ebene: 3 - Daten verifizieren

C1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze, die Qualität bzw. Plausibilität von Daten zu prüfen?

1) Ich übernehme die Daten unkritisch. Ich habe kein Bewusstsein dafür, wie und warum Daten kritisch zu bewerten sind. ☐

2) Ich führe eine Überprüfung anhand einfacher Datenqualitätskriterien (z. B. Vollständigkeit, keine Duplikate) durch. ☐

3) Ich überprüfe die Daten in mehreren Schritten. Dafür wende ich Standardprozeduren an. Datenquellen werden überprüft, verschiedene Quellen berücksichtigt und kritisch beurteilt. ☐

4) Ich orientiere mich an den bestehenden Best Practices des KWB (z. B. festgeschrieben im Qualitätsmanagement). Ich wurde in den Methoden des Datenhandlings geschult. Datenquellen werden von mir transparent verwendet und bieten Zugriff auf die Originalquellen, damit Dritte sie selbst überprüfen können. ☐

C2. SOLL-Zustand: Daten verifizieren

1) Ich übernehme die Daten unkritisch. Ich habe kein Bewusstsein dafür, wie und warum Daten kritisch zu bewerten sind. ☐

2) Ich führe eine Überprüfung anhand einfacher Datenqualitätskriterien (z. B. Vollständigkeit, keine Duplikate) durch. ☐

3) Ich überprüfe die Daten in mehreren Schritten. Dafür wende ich Standardprozeduren an. Datenquellen werden überprüft, verschiedene Quellen berücksichtigt und kritisch beurteilt. ☐

4) Ich orientiere mich an den bestehenden Best Practices des KWB (z. B. festgeschrieben im Qualitätsmanagement). Ich wurde in den Methoden des Datenhandlings geschult. Datenquellen werden von mir transparent verwendet und bieten Zugriff auf die Originalquellen, damit Dritte sie selbst überprüfen können. ☐

Teil D: Individuelle Ebene: 4 - Daten säubern

D1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze Daten zu „bereinigen“?

1) Ich habe kein Bewusstsein dafür, dass vorliegende Daten überprüft und bereinigt werden müssen. Daten werden so weiterverwendet, wie sie zur Verfügung gestellt wurden. ☐

2) Ich habe ein Bewusstsein dafür, dass Daten nicht immer systematisch formatiert sind (z. B. kann eine Excel-Spalte Text, Datum und numerische Werte enthalten). Die Datenbereinigung wird von mir manuell in Excel vorgenommen. Ich habe eigene grundlegende Kriterien zur Datenqualität definiert und kann diese anwenden (z. B. einfacher Umgang mit Werten unter der Bestimmungsgrenze). ☐

3) Ich habe ein großes Bewusstsein für Datenqualitätskriterien. Ich verwende existierende Software-Tools zur Unterstützung bei der Datenreinigung (z. B. click-basierte Tools mit graphischer Oberfläche: Open Refine). Damit kann ich beispielsweise String-Manipulationen (Suchen und Ersetzen) durchführen, um sicherzustellen, dass genormte Bezeichnungen einheitlich verwendet werden. ☐

4) Ich habe ein großes Bewusstsein für Datenqualitätskriterien. Ich verwende selbst programmierte Skripte (z. B. in der Software „R“) für die Datenreinigung. Das Vorgehen ist im Quellcode dokumentiert und kann durch Ausführen der Skripte reproduziert werden. ☐

D2. SOLL-Zustand: Daten säubern

1) Ich habe kein Bewusstsein dafür, dass vorliegende Daten überprüft und bereinigt werden müssen. Daten werden so weiterverwendet, wie sie zur Verfügung gestellt wurden. ☐

2) Ich habe ein Bewusstsein dafür, dass Daten nicht immer systematisch formatiert sind (z. B. kann eine Excel-Spalte Text, Datum und numerische Werte enthalten). Die Datenbereinigung wird von mir manuell in Excel vorgenommen. Ich habe eigene grundlegende Kriterien zur Datenqualität definiert und kann diese anwenden (z. B. einfacher Umgang mit Werten unter der Bestimmungsgrenze). ☐

3) Ich habe ein großes Bewusstsein für Datenqualitätskriterien. Ich verwende existierende Software-Tools zur Unterstützung bei der Datenreinigung (z. B. click-basierte Tools mit graphischer Oberfläche: Open Refine). Damit kann ich beispielsweise String-Manipulationen (Suchen und Ersetzen) durchführen, um sicherzustellen, dass genormte Bezeichnungen einheitlich verwendet werden. ☐

4) Ich habe ein großes Bewusstsein für Datenqualitätskriterien. Ich verwende selbst programmierte Skripte (z. B. in der Software „R“) für die Datenreinigung. Das Vorgehen ist im Quellcode dokumentiert und kann durch Ausführen der Skripte reproduziert werden. ☐

Teil E: Individuelle Ebene: 5 - Daten analysieren

E1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze bei der Datenanalyse?

1) Ich nutze Excel, um einfache statistische Werte zu ermitteln (Mittelwert, Minimum, Maximum, usw.) ☐

2) Ich nutze Excel, um weitergehende Analysen (deskriptive Statistik) durchzuführen (z. B. mit Hilfe von Pivot-Tabellen, Histogrammen, Boxplots oder dergleichen). ☐

3) Ich kann Korrelationen zwischen Variablen in einem Datensatz - unter Berücksichtigung von statistischen Unsicherheiten - herstellen. Hierbei verwende ich z. B. Excel oder Origin. ☐

4) Ich nutze Programmiersprachen (wie z. B. "R") zur statistischen Auswertung und dem Erstellen von Prognosen und Vorhersagen. Ggf. umfasst meine Analyse auch die Anwendung von numerischen Simulationsmodellen (z. B. EPANET, MODFLOW) ☐

E2. SOLL-Zustand: Daten analysieren

Ich nutze Excel, um einfache statistische Werte zu ermitteln (Mittelwert, Minimum, Maximum, usw.) ☐

Ich nutze Excel, um weitergehende Analysen (deskriptive Statistik) durchzuführen (z. B. mit Hilfe von Pivot-Tabellen, Histogrammen, Boxplots oder dergleichen). ☐

Ich kann Korrelationen zwischen Variablen in einem Datensatz - unter Berücksichtigung von statistischen Unsicherheiten - herstellen. Hierbei verwende ich z. B. Excel oder Origin. ☐

Ich nutze Programmiersprachen (wie z. B. "R") zur statistischen Auswertung und dem Erstellen von Prognosen und Vorhersagen. Ggf. umfasst meine Analyse auch die Anwendung von numerischen Simulationsmodellen (z. B. EPANET, MODFLOW) ☐

Teil F: Individuelle Ebene: 6 - Daten visualisieren

F1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze bei der Datenvisualisierung?

1) Ich erstelle einfache Visualisierungen (z. B. Balken- oder Kuchendiagramme) in Excel. Mir fehlt das Wissen, in welchem Kontext welche Visualisierung idealerweise verwendet werden sollte. Meine Entscheidung basiert auf dem Bauchgefühl oder danach, was gut aussieht (trial and error). ☐

2) Ich erstelle komplizierte Visualisierungen (Pivot-Tabellen, Histogramme, Boxplots), die auf zum Teil umfangreichen Analysen beruhen. Dazu verwende ich Excel oder Origin. ☐

3) Ich erstelle interaktive Visualisierungen, wobei Unsicherheiten in den Daten immer mitkommuniziert werden. Dabei nutze ich grafische Software-Tools (click-basiert: wie z. B. Tableau, CartoDB) ☐

4) Ich verwende vielfältige Tools zur Visualisierung, die an die Benutzeranforderungen anpassbar sind. Diese werden von mir selbst programmiert (z. B. interaktive Visualisierungen in R mit dem Paket „shiny"). ☐

F2. SOLL-Zustand: Daten visualisieren

1) Ich erstelle einfache Visualisierungen (z. B. Balken- oder Kuchendiagramme) in Excel. Mir fehlt das Wissen, in welchem Kontext welche Visualisierung idealerweise verwendet werden sollte. Meine Entscheidung basiert auf dem Bauchgefühl oder danach, was gut aussieht (trial and error). ☐

2) Ich erstelle komplizierte Visualisierungen (Pivot-Tabellen, Histogramme, Boxplots), die auf zum Teil umfangreichen Analysen beruhen. Dazu verwende ich Excel oder Origin. ☐

3) Ich erstelle interaktive Visualisierungen, wobei Unsicherheiten in den Daten immer mitkommuniziert werden. Dabei nutze ich grafische Software-Tools (click-basiert: wie z. B. Tableau, CartoDB) ☐

4) Ich verwende vielfältige Tools zur Visualisierung, die an die Benutzeranforderungen anpassbar sind. Diese werden von mir selbst programmiert (z. B. interaktive Visualisierungen in R mit dem Paket „shiny"). ☐

Teil G: Individuelle Ebene: 7 - Daten kommunizieren

G1. IST-Zustand: Welche Aussage beschreibt am besten deine Ansätze, die Ergebnisse deiner datenbezogenen Arbeit zu kommunizieren?

1) Meine Erkenntnisse aus den Daten kommuniziere ich nicht. ☐

2) Ich nutze statische Visualisierungen für meine Texte (z.B. in Berichten, Veröffentlichungen, Präsentationen, Website). ☐

3) In meinen Projekten werden Ergebnisse über verschiedene Kanäle (z.B. Vorträge, Blogpost, Newsmeldungen) kommuniziert. ☐

4) In meinen Projekten werden Ergebnisse über verschiedene Kanäle (z.B. Vorträge, Blogpost, Newsmeldungen) kommuniziert. Zudem sind die Datengrundlagen und Methoden in öffentlichen Repositorien zugänglich. ☐

G2. SOLL-Zustand: Daten kommunizieren?

1) Meine Erkenntnisse aus den Daten kommuniziere ich nicht. ☐

2) Ich nutze statische Visualisierungen für meine Texte (z.B. in Berichten, Veröffentlichungen, Präsentationen, Website) ☐

3) In meinen Projekten werden Ergebnisse über verschiedene Kanäle (z.B. Vorträge, Blogpost, Newsmeldungen) kommuniziert. ☐

4) In meinen Projekten werden Ergebnisse über verschiedene Kanäle (z.B. Vorträge, Blogpost, Newsmeldungen) kommuniziert. Zudem sind die Datengrundlagen und Methoden in öffentlichen Repositorien zugänglich. ☐

Teil H: Organisationsebene: 1 - Daten-Kultur

H1. IST-Zustand: Welche Aussage beschreibt am besten die Kultur am KWB hinsichtlich der Arbeit mit Daten?

1) Es gibt keine Mitarbeiter/innen mit datenbezogenem Hintergrund. ☐

2) Wissenschaftliche Mitarbeiter/innen (Projektmitarbeiter, Projektleiter, Abteilungsleiter) mit datenbezogenem Hintergrund sind am KWB vorhanden. ☐

3) Wissenschaftliche Mitarbeiter/innen (Projektmitarbeiter, Projektleiter, Abteilungsleiter) mit datenbezogenem Hintergrund sind am KWB vorhanden. Auch die Geschäftsführung ist sich der Notwendigkeit bewusst, Daten zu sammeln, zu speichern und zielgerichtet einzusetzen. Datenanalyse ist Bestandteil der täglichen Arbeit. ☐

4) Das „Datenhandling“ (vom Umgang mit Rohdaten über die Analyse und Interpretation bis hin zur Kommunikation) ist im Qualitätsmanagement verankert und wird umgesetzt. Datenkompetenz ist eine wichtige Qualifikation für die meisten Mitarbeiter/innen. Ausreichend Ressourcen (Zeit, Budget, Personal) und Weiterbildungsmöglichkeiten für das Arbeiten mit Daten existieren. ☐

H2. SOLL-Zustand: Daten-Kultur

1) Es gibt keine Mitarbeiter/innen mit datenbezogenem Hintergrund. ☐

2) Wissenschaftliche Mitarbeiter/innen (Projektmitarbeiter, Projektleiter, Abteilungsleiter) mit datenbezogenem Hintergrund sind am KWB vorhanden. ☐

3) Wissenschaftliche Mitarbeiter/innen (Projektmitarbeiter, Projektleiter, Abteilungsleiter) mit datenbezogenem Hintergrund sind am KWB vorhanden. Auch die Geschäftsführung ist sich der Notwendigkeit bewusst, Daten zu sammeln, zu speichern und zielgerichtet einzusetzen. Datenanalyse ist Bestandteil der täglichen Arbeit. ☐

4) Das „Datenhandling" (vom Umgang mit Rohdaten über die Analyse und Interpretation bis hin zur Kommunikation) ist im Qualitätsmanagement verankert und wird umgesetzt. Datenkompetenz ist eine wichtige Qualifikation für die meisten Mitarbeiter/innen. Ausreichend Ressourcen (Zeit, Budget, Personal) und Weiterbildungsmöglichkeiten für das Arbeiten mit Daten existieren. ☐

Teil I: Organisationsebene: 2 - Urheberrecht und Datenschutz

I1. IST-Zustand: Welche Aussage beschreibt am besten den Umgang am KWB mit Datenschutz und Urheberrecht?

1) Mit Daten wird ohne Rücksicht auf Datenschutz und Urheberrecht gearbeitet. Es sind keine Richtlinien definiert, die die Vertraulichkeit, Integrität und Verfügbarkeit der Daten gewährleisten. ☐

2) Einige Mitarbeiter/innen sind sich möglicher Datenschutzprobleme bewusst. Es gibt Bestrebungen, Daten sicher zu behandeln, aber es existieren keine internen Richtlinien zum Datenschutz. ☐

3) Datenschutzrichtlinien und ein verantwortlicher Umgang mit Daten sind definiert. Organisationsweite Richtlinien für eine sichere Datenverarbeitung unter Berücksichtigung des Urheberrechts (z. B. Nutzungsbeschränkungen durch Dateneigentümer für ein bestimmtes Projekt) existieren. ☐

4) Organisationsweite Richtlinien zum Datenschutz, der Datenverarbeitung und dem Urheberrecht existieren. Diese werden laufend aktualisiert und konsequent umgesetzt. ☐

I2. SOLL-Zustand: Urheberrecht und Datenschutz

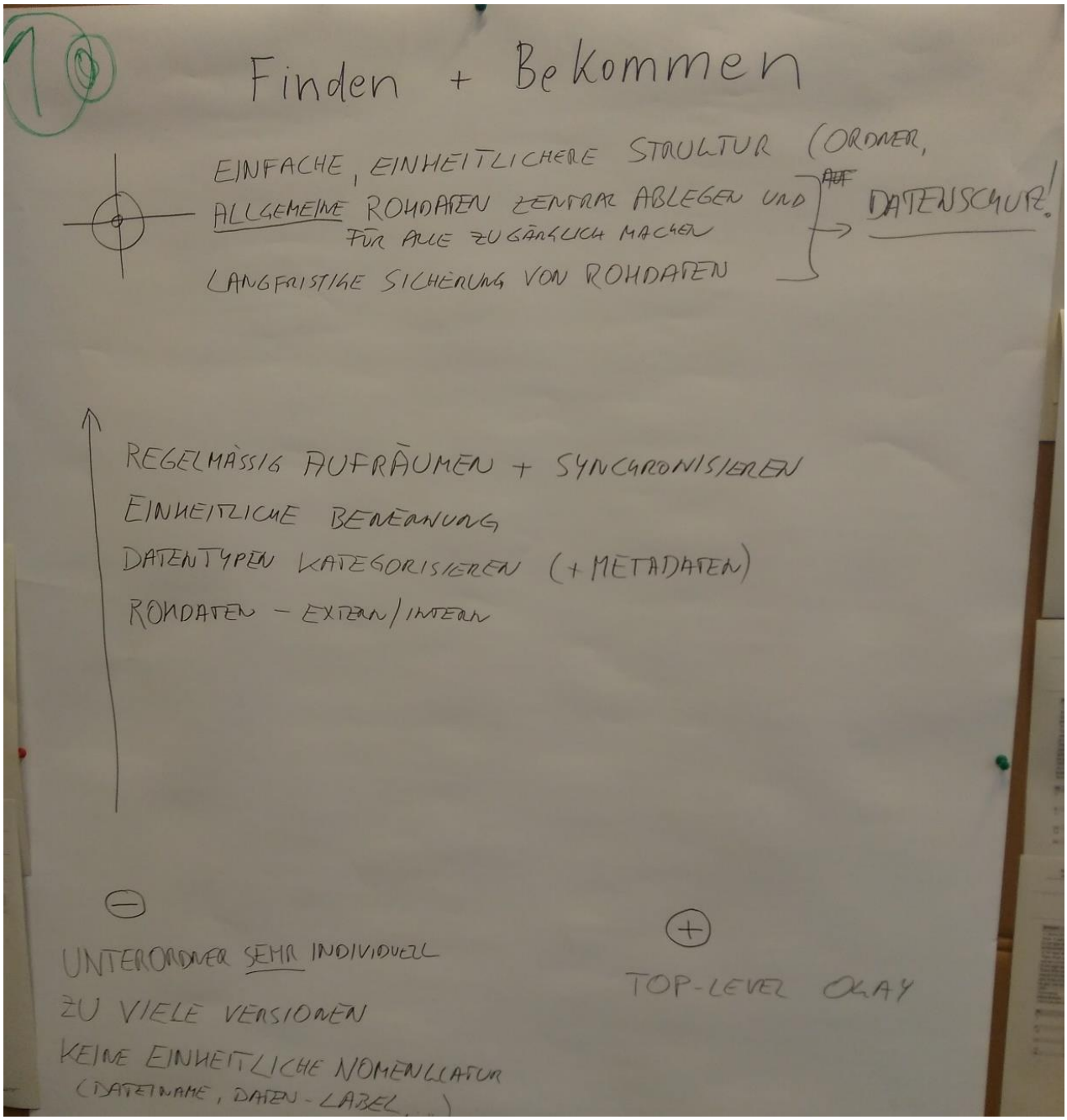
1) Mit Daten wird ohne Rücksicht auf Datenschutz und Urheberrecht gearbeitet. Es sind keine Richtlinien definiert, die die Vertraulichkeit, Integrität und Verfügbarkeit der Daten gewährleisten. ☐

2) Einige Mitarbeiter/innen sind sich möglicher Datenschutzprobleme bewusst. Es gibt Bestrebungen, Daten sicher zu behandeln, aber es existieren keine internen Richtlinien zum Datenschutz. ☐

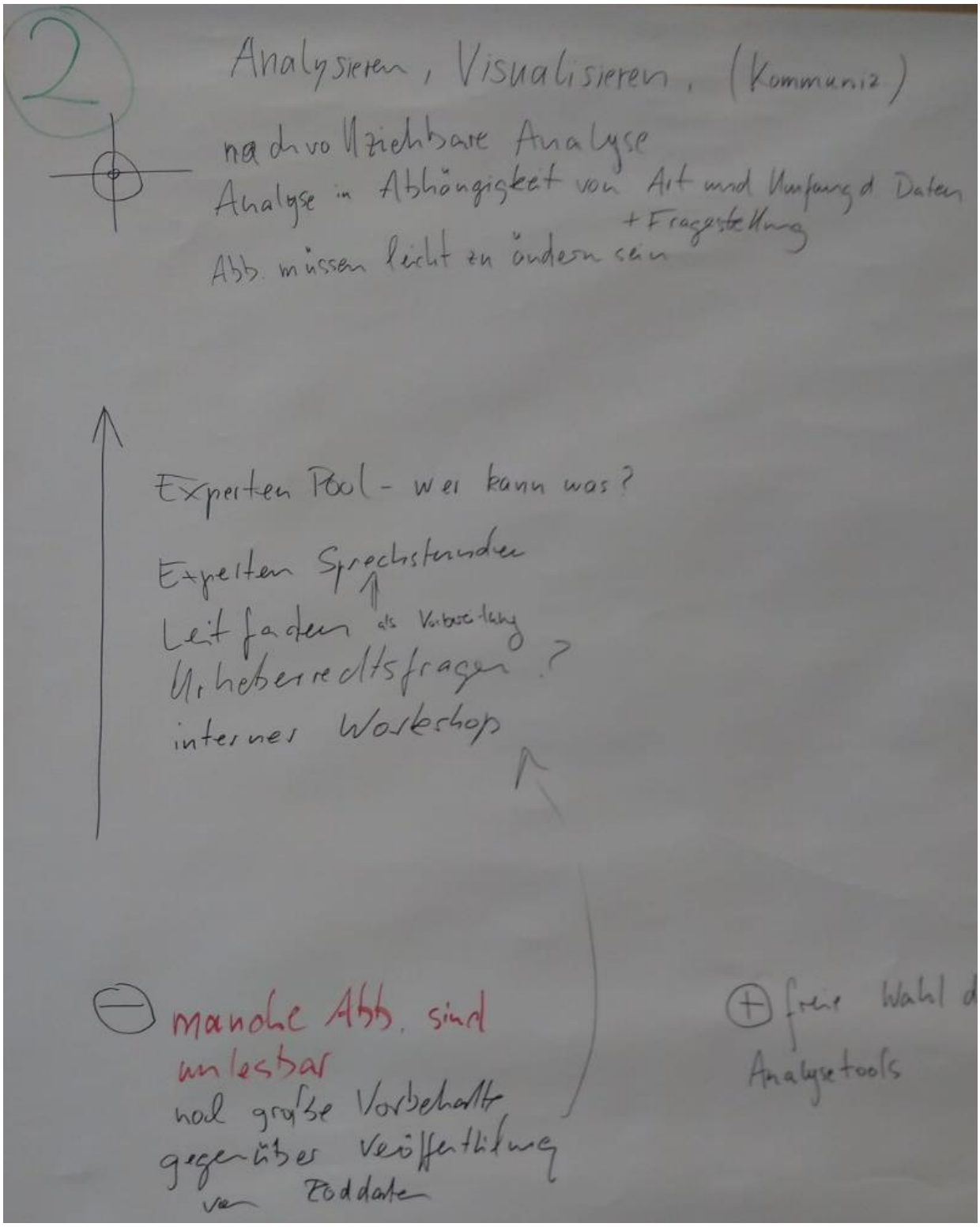
3) Datenschutzrichtlinien und ein verantwortlicher Umgang mit Daten sind definiert. Organisationsweite Richtlinien für eine sichere Datenverarbeitung unter Berücksichtigung des Urheberrechts (z. B. Nutzungsbeschränkungen durch Dateneigentümer für ein bestimmtes Projekt) existieren. ☐

4) Organisationsweite Richtlinien zum Datenschutz, der Datenverarbeitung und dem Urheberrecht existieren. Diese werden laufend aktualisiert und konsequent umgesetzt. ☐

Anhang D: **Brainstorming zum Themenkomplex „Daten finden und bekommen“**



Anhang E: **Brainstorming zum Themenkomplex „Daten analysieren, visualisiern und kommunizieren“**

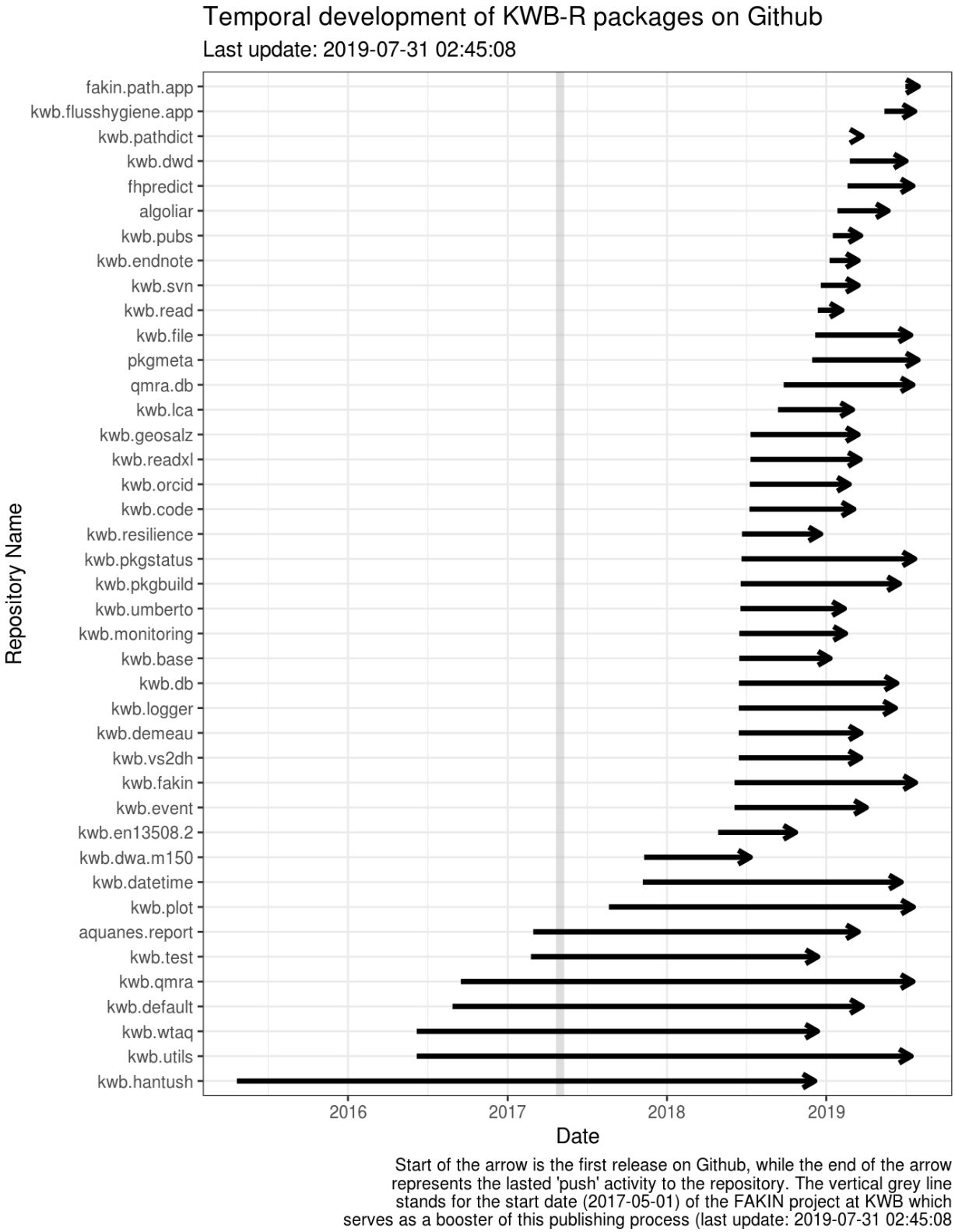


Anhang F: **Teilnehmerliste FAKIN Abschlussworkshop**

Teilnehmerliste FAKIN Abschlussworkshop (25.07.2019, 10 – 13 Uhr, Kompetenzzentrum Wasser Berlin gGmbH, Cicerostraße 24, 10709 Berlin)

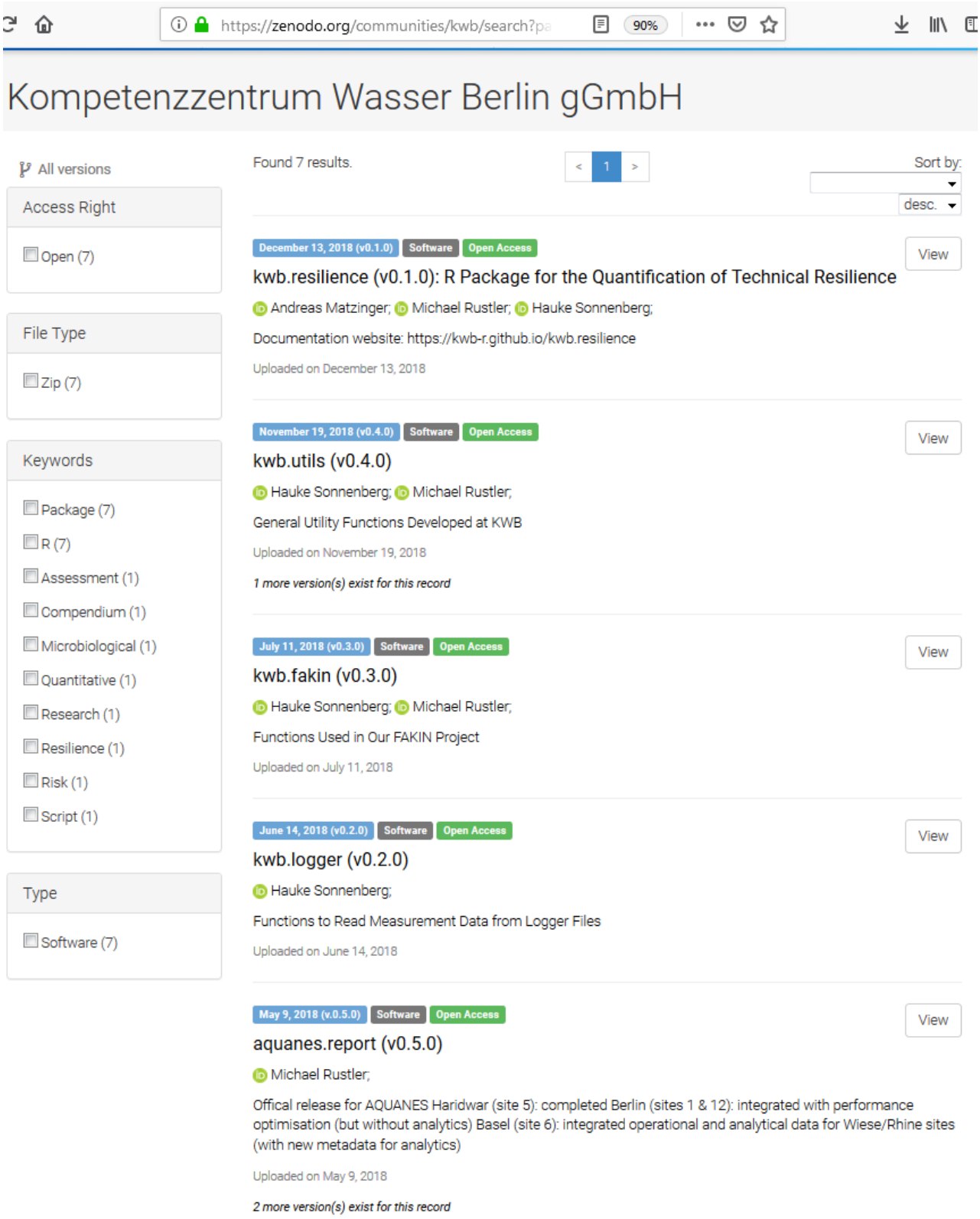
Institut	Name	Unterschrift
3S Consult GmbH	Ingo Kropp	
adelphi research gGmbH	Anika Conrad	
Blue Biolabs GmbH	Oliver Thronicker	
Freie Universität Berlin	Dr. Andreas Winkler	
Ifak - Institut für Automation und Kommunikation e. V.	Prof. Ulrich Jumar	
Kompetenzzentrum Wasser Berlin gGmbH	Dr. Hella Schwarzmüller	
Kompetenzzentrum Wasser Berlin gGmbH	Hauke Sonnenberg	
Kompetenzzentrum Wasser Berlin gGmbH	Michael Rustler	
Papiertechnische Stiftung	Dr. Tiemo Arndt	
Pecher und Partner	Philipp Birkholz	
Reiner Lemoine Institut gGmbH	Christian Hofmann	
Sächsisches Institut für die Druckindustrie GmbH	Beatrix Genest	
Sächsisches Textilforschungsinstitut e.V.	Dr. Heike Illing-Günther	
Textilforschungsinstitut Thüringen-Vogtland e.V	Sabine Gimpel	
Uni Kaiserslautern	Jonas Neumann	
Papiertechnische Stiftung PTS	Maria Tillen	
Dr. Schulmacher - Ingenieurbüro f. Wasser u. Umwelt	Dominik Kolesch Patrick Stolz	

Anhang G: **Veröffentlichung von Programmcode (R Pakete) auf GitHub (https://kwb-r.github.io/status)**



Reference: *https://github.com/KWB-R/pkgmeta/blob/8353c358c2cb3bc49bfb1977a767031c3053a5ff/articles/visualisation_files/figure-html/unnamed-chunk-1-1.png*

Anhang H: Veröffentlichungen von Programmcode (R Pakete) auf Zenodo (https://zenodo.org/communities/kwb)



IV. LITERATUR

AirBnB (2018) A next-generation curated knowledge sharing platform for data scientists and other technical professions (v0.8.8), v0.8.8, GitHub, https://github.com/airbnb/knowledge-repo/releases/tag/v0.8.8 (abgerufen am: 2019-12-19)

AirbnbEng (2016) Scaling Knowledge at Airbnb (Airbnb Engineering & Data Science), Blogpost, https://medium.com/airbnb-engineering/scaling-knowledge-at-airbnb-875d73eff091 (abgerufen am: 2019-12-19)

Appveyor (2020) Continous Integration and Deployment Service for Windows, Linux and macOS. https://www.appveyor.com/ (abgerufen am: 2020-01-08)

Bertelmann, R., Gebauer, P., Hasler, T., Kirchner, I., Peters-Kottig, W., Razum, M., Recker, A., Ulbricht , D. & van Gasselt, S. (2014) Einstieg ins Forschungsdatenmanagement in den Geowissenschaften. Deutsches GeoForschungsZentrum doi:10.2312/lis.14.01 (abgerufen am: 2019-12-19)

Bion, R., Chang, R. & Goodman, J. (2018) How R Helps Airbnb Make the Most of its Data. *The American Statistician* 72 (1)**:** 46-52 doi:10.1080/00031305.2017.1392362

Boettiger, C., Salmon, M., Sonnenberg, H., Ross, N., Leinweber, K., Smith, A., Meyer, S., Rustler, M., Woo, K. & Krystalli, A. (2019) codemetar (v.0.1.7): Generate CodeMeta Metadata for R Packages, v.0.1.7, Zenodo, doi:10.5281/zenodo.2598516 (abgerufen am: 2019-12-19)

Codecov (2020) Code Coverage Done Right. https://codecov.io/ (abgerufen am: 2020-01-08)

CodeMeta (2019) The CodeMeta Project. https://codemeta.github.io (abgerufen am: 2019-12-19)

DataCarpentry (2019) Data Organization in Spreadsheets for Ecologists - Formatting data tables in Spreadsheets. https://datacarpentry.org/spreadsheet-ecology-lesson/01-format-data/ (abgerufen am: 2019-12-19)

DataOne (2019) Best Practices database. https://www.dataone.org/best-practices (abgerufen am: 2019-12-19)

Datenschule (2017) Evaluation des Data Literacy Maturity Models. https://datenschule.de/files/workshops/DateLiteracyModel-Matrix-Datenschule.pdf (abgerufen am: 2019-12-19)

EC (2017) H2020 Programme - Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 (v3.2). European Commission https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf (abgerufen am: 2019-12-19)

FDMentor (2019) Projektergebnisse aus dem Verbundprojekt „Erarbeitung generalisierbarer Strategien und Lösungen für das Forschungsdatenmanagement unter Einbeziehung bestehender Expertise an universitären Zentraleinrichtungen. https://zenodo.org/communities/fdmentor (abgerufen am: 2019-12-18)

Git (2020) Free and Open Source Distributed Version Control System. https://git-scm.com/ (abgerufen am: 2020-01-08)

GitHub (2020) The world`s leading software development platform. https://github.com/ (abgerufen am: 2020-01-08)

GitHub Pages (2020) Websites for you and your projects. https://pages.github.com/ (abgerufen am: 2020-01-08)

Hartmann, N. K., Jacob, B. & Weiß, N. (2019) RISE-DE - Referenzmodell für Strategieprozesse im institutionellen Forschungsdatenmanagement (v1.0). Universität Potsdam doi:10.5281/zenodo.3516620 (abgerufen am: 2019-12-19)

Helbig, K., Biernacka, K., Buchholz, P., Dolzycka, D., Hartmann, N., Hartmann, T., Hiemenz, B. M., Jacob, B., Kuberek, M., Weiß, N. & Dreyer, M. (2019) Lösungen und Leitfäden für das institutionelle Forschungsdatenmanagement. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 6 (3) doi:10.5282/o-bib/2019H3S21-39

Hiemenz, B. & Kuberek, M. (2018a) Empfehlungen zur Erstellung institutioneller Forschungsdaten-Policies. Das Forschungsdaten-Policy-Kit als generischer Baukasten mit Leitfragen und Textbausteinen für Hochschulen in Deutschland. TU Berlin doi:10.14279/depositonce-7521 (abgerufen am: 2019-12-19)

Hiemenz, B. M. & Kuberek, M. (2018b) Leitlinie? Grundsätze? Policy? Richtlinie? – Forschungsdaten-Policies an deutschen Universitäten. *o-bib. Das offene Bibliotheksjournal / Herausgeber VDB* 5 (2) doi:10.5282/o-bib/2018H2S1-13

Hugo (2019) The fastest framework for building websites. https://gohugo.io (abgerufen am: 2019-12-19)

ISO-8601 (2019) 8601-1:2019: Date and time - Representations for information interchange, International Organization for Standardization, https://www.iso.org/obp/ui#iso:std:iso:8601:-1:ed-1:v1:en (abgerufen am: 2019-12-19)

KWB (2019) Qualitätsmanagement: Wasserforschung mit Qualität - unsere Qualitätspolitik. https://www.kompetenz-wasser.de/de/quality/ (abgerufen am: 2019-12-19)

LimeSurvey (2020) Professionelle Online-Umfragen mit LimeSurvey. https://www.limesurvey.org/de/ (abgerufen am: 2020-01-08)

MyBinder (2020) Turn a Git repo into a collection of interactive notebooks. https://mybinder.org/ (abgerufen am: 2020-01-08)

ODM-2 (2020) Observations Data Model 2 - An information model and supporting software ecosystem for feature-based Earth observations. http://www.odm2.org/ (abgerufen am: 2020-01-08)

OptimalWorkshop (2019) Treejack - Information architecture validation software. https://www.optimalworkshop.com/treejack (abgerufen am: 2019-12-19)

R Core Team (2020) A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.r-project.org (abgerufen am: 2020-01-08)

Rustler, M. (2016a) kwb.qmra (v.0.1.1): An R package for QMRA (quantitative microbial risk assessment) of water supply systems, v0.1.1, Zenodo, doi:10.5281/zenodo.154111 (abgerufen am:

Rustler, M. (2016b) Quantitative microbiological risk assessment for different wastewater reuse options in Old Ford (v.1.0), v1.0, Zenodo, doi:10.5281/zenodo.159527 (abgerufen am: 2019-12-19)

Rustler, M. (2019a) FAKIN Wissensspeicher – ein proof-of-concept Beispiel für einen institutionellen Wissensspeicher. https://kwb-r.github.io/fakin (abgerufen am: 2019-12-19)

Rustler, M. (2019b) Lessons-learnt: Auswertung der FAKIN Wissensspeicher Umfrage Blogpost, https://kwb-r.github.io/fakin/post/workshop-lessons-learnt/ (abgerufen am: 2019-12-19)

Rustler, M. (2020a) kwb.pkgstatus (v0.1.0): R Package for Checking KWB Package Status v0.1.0, Zenodo, doi:10.5281/zenodo.3604014 (abgerufen am: 2020-01-10)

Rustler, M. (2020b) pkgmeta (v0.1.0): R Package for Meta-Analysis of KWB-R Packages on GitHub, v0.1.0, Zenodo, doi:10.5281/zenodo.3604093 (abgerufen am: 2020-01-10)

Rustler, M. & Sonnenberg, H. (2019) kwb.pkgbuild (v0.1.2): R Package for Standardised Development at KWB, v0.1.2, Zenodo, doi:10.5281/zenodo.3430683 (abgerufen am: 2019-12-19)

Rustler, M. & Sonnenberg, H. (2020) kwb.umberto (v0.1.0): R Package for Supporting LCA with UMBERTO Software at KWB, v0.1.0, Zenodo, doi:10.5281/zenodo.3604007 (abgerufen am: 2020-01-10)

Rustler, M., Sonnenberg, H. & Sprenger, C. (2019a) Best Practices in Research Data Management. Kompetenzzentrum Wasser Berlin gGmbH https://kwb-r.github.io/fakin.doc/ (abgerufen am: 2019-12-19)

Rustler, M., Sonnenberg, H. & Sprenger, C. (2019b) Creating a Knowledge Repo for a Small Research Institute. https://kwb-r.github.io/fakin/abstract/2019-06-04_derse19/abstract/knowledge-repo.pdf (abgerufen am: 2019-12-19)

Sonnenberg, H. (2016) kwb.wtaq (v0.2.1): An R package for interfacing well drawdown model WTAQ, v0.2.1, Zenodo, doi:10.5281/zenodo.61610 (abgerufen am: 2019-12-19)

Sonnenberg, H. (2019) kwb.datetime (v0.4.0): R Package for Date Time Handling, v0.4.0, Zenodo, doi:10.5281/zenodo.3484082 (abgerufen am: 2019-12-19)

Sonnenberg, H. & Rustler, M. (2020) fakin.path.app (v0.3.0): R Shiny Application for Path Analysis, v.0.3.0, Zenodo, doi:10.5281/zenodo.3603503 (abgerufen am: 2020-01-10)

Sonnenberg, H., Rustler, M., Riechel, M., Caradot, N., Matzinger, A. & Rouault, P. (2013) Best data handling practices in water-related research. *Water Practice & Technology* Vol 8 (No 3-4)**:** 390-398 doi:10.2166/wpt.2013.039

Sternkopf (2017) Doing Good with Data - Development of a Maturity Model for Data Literacy in Non-Governmental Organizations. Master thesis. Faculty of Business and Economics, Hochschule für Wirtschaft und Recht Berlin https://datenschule.de/files/workshops/DoingGoodwithData-DevelopmentofaMaturityModelforDataLiteracyinNGOs_HelenaSternkopf.pdf (abgerufen am: 2019-12-19)

Subversion (2020) Enterprise-class centralized version control for the masses. https://subversion.apache.org/ (abgerufen am: 2020-01-08)

Travis-CI (2020) The simplest way to test and deploy your projects. https://travis-ci.com/ (abgerufen am: 2020-01-08)

Wikipedia (2020a) Programmcode. https://de.wikipedia.org/wiki/Programmcode (abgerufen am: 2020-01-08)

Wikipedia (2020b) Repository. https://de.wikipedia.org/wiki/Repository (abgerufen am: 2020-01-08)

Xie, Y. (2019) blogdown: Create Blogs and Websites with R Markdown, R package version 0.17, https://cran.r-project.org/web/packages/blogdown (abgerufen am: 2019-12-19)

Zenodo (2020) Research - Shared. https://zenodo.org/ (abgerufen am: 2020-01-08)