

FAKIN Abschluss-Workshop

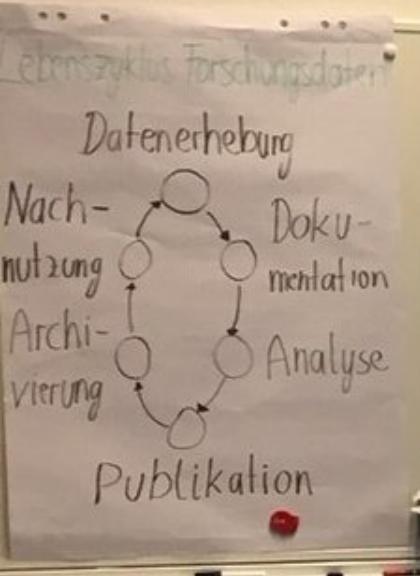


By Jørgen Stamp,
<http://digitalbevaring.dk/>

Gefördert durch:



Bundesministerium
für Bildung
und Forschung



Agenda:

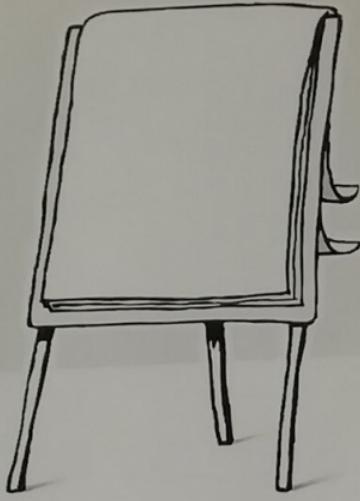
Vorstellungsrunde
Unsere Datentreise
Pause (10:15-10:30)
Diskussion
Frage und Antwort



A light-colored wooden conference table is set up in the foreground. On the table, there is a laptop, a red folder, a bottle of water, a glass of water, a small white container, a red pen, and some papers. A red office chair is partially visible under the table on the left. In the background, a white curved sofa is positioned against a wall with vertical blinds.

Agenda:

- Vorstellungsrunde
- Unsere Datenreise
Pause (11¹⁵ - 30)
- Tools
- Diskussion
- „Get Together“ Lunch (13⁰⁰ - ?)



Abschluss Workshop

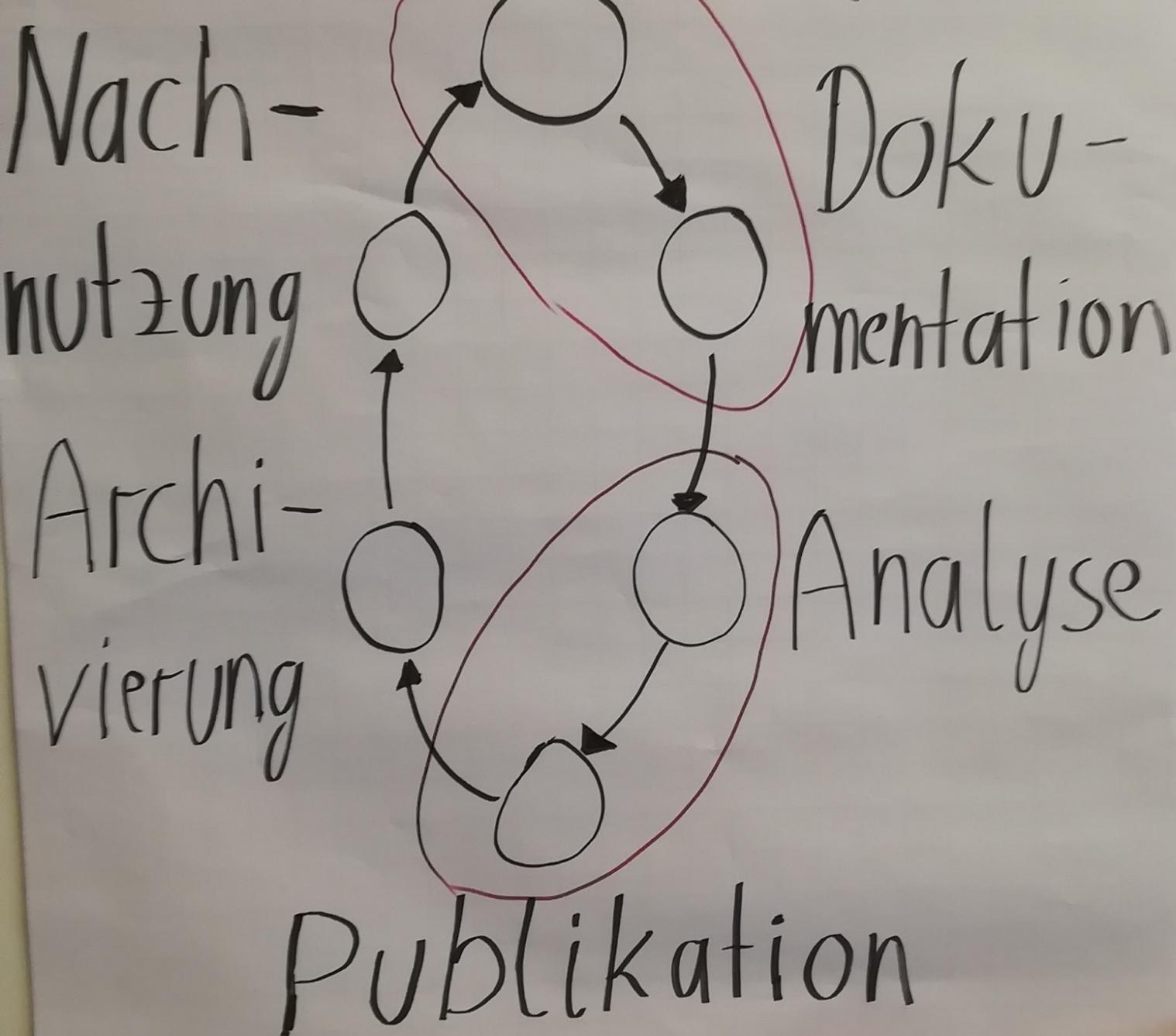
Juli 2019



- 1. Unser Vorgehen**
 - 2. Unsere Ergebnisse**
 - 3. Ihre FDM**
- Herausforderungen
(und Lösungen)?**

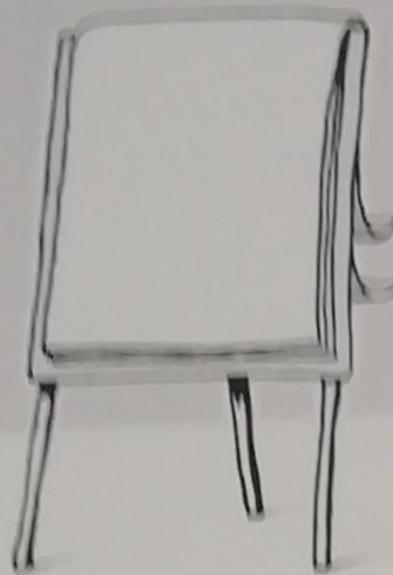
Lebenszyklus Forschungsdaten

Datenerhebung



Auftaktworkshop

September 2017



1. Finden

und

Bekommen



2. Analysieren,

Visualisieren

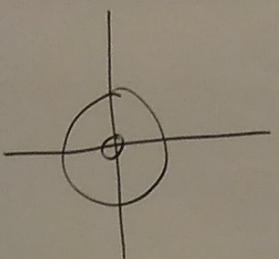
und

Kommunizieren



10

Finden + Bekommen



EINFACHE, EINHEITLICHERE STRUKTUR (ORDNER,
ALLGEMEINE ROHDATEN ZENTRAL ABLEGEN UND
FÜR ALLE ZUGÄNGLICH MACHEN
LANGFRISTIGE SICHERUNG VON ROHDATEN

→ DATENSICHERHEIT!

REGELMÄSSIG AUFRÄUMEN + SYNCHRONISIEREN
EINHEITLICHE BEZEICHNUNG
DATENTYPEN KATEGORISIEREN (+ METADATEN)
ROHDATEN - EXTERN / INTERN

(-)

UNTERORDNER SEHR INDIVIDUELL

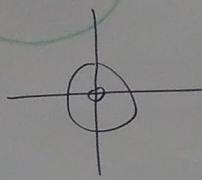
ZU VIELE VERSIONEN

KEINE EINHEITLICHE NOMENKLATUR
(DATENAME, DATEN-LABEL,...)

(+)

TOP-LEVEL OKAY

2



Analysieren, Visualisieren, (Kommuniz.)

nachvollziehbare Analyse

Analyse in Abhängigkeit von Art und Umfang d. Daten

Abb. müssen leicht zu ändern sein + Fragestellung



Experten Pool - Wer kann was?

Experten Sprechstunden

Leitfäden als Vorberatung

Urheberrechtsfragen?

internes Workshop



manche Abb. sind
unlesbar
noch große Vorbehalte
gegenüber Veröffentlichung
von Podcasts

(+) freie Wahl der
Analysetools



Best-practices Workshop

Januar 2018

- Rohdaten



- Ordnerstruktur



DIN-Norm?

- Nomenklatur



„überholt“-Ordner

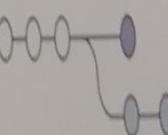
- Metadaten



Data package

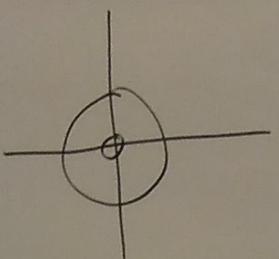
eigener Standard

- Versionierung



10

Finden + Bekommen



EINFACHE, EINHEITLICHERE STRUKTUR (ORDNER,
ALLGEMEINE ROHDATEN ZENTRAL ABLEGEN UND
FÜR ALLE ZUGÄNGLICH MACHEN
LANGFRISTIGE SICHERUNG VON ROHDATEN

AFF DATENSCHUTZ!



REGELMÄSSIG AUFRÄUMEN + SYNCHRONISIEREN
EINHEITLICHE BEZEICHNUNG
DATENTYPEN KATEGORISIEREN (+ METADATEN)
ROHDATEN - EXTERN / INTERN

(-)

UNTERORDNER SEMI INDIVIDUELL

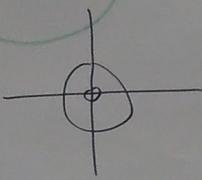
ZU VIELE VERSIONEN

KEINE EINHEITLICHE NOMENKLATUR
(DATENAME, DATEN-LABEL,...)

(+)

TOP-LEVEL OKAY

2



Analysieren, Visualisieren, (Kommuniz.)

nachvollziehbare Analyse

Analyse in Abhängigkeit von Art und Umfang d. Daten

Abb. müssen leicht zu ändern sein + Fragestellung

Experten Pool - Wer kann was?

Experten Sprechstunden

Leitfäden als Vorbehaltung
Urheberrechtsfragen?

internes Workshop



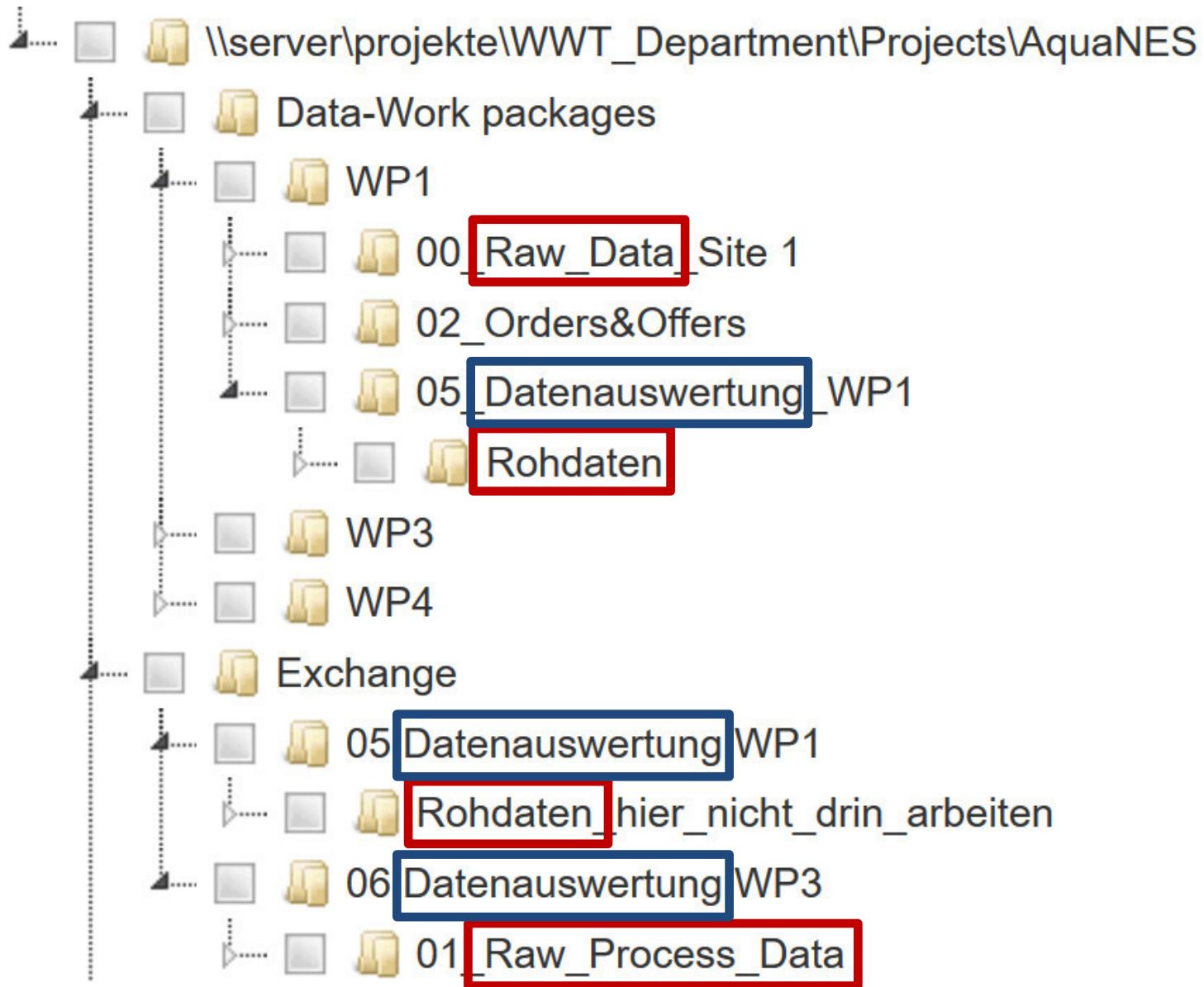
manche Abb. sind
unlesbar
noch große Vorbehalte
gegenüber Veröffentlichung
von Podcasts

(+) freie Wahl der
Analysetools

\\server\projekte\WWT_Department\Projects\AquaNES

- Administration
- Closure & Evaluation
- Communication
- Data-Work packages
 - offene Dokumente
 - WP1
 - WP3
 - WP4
 - WP5
 - WP6
 - WP7
 - WP8
 - WP9
- Exchange
- Reports

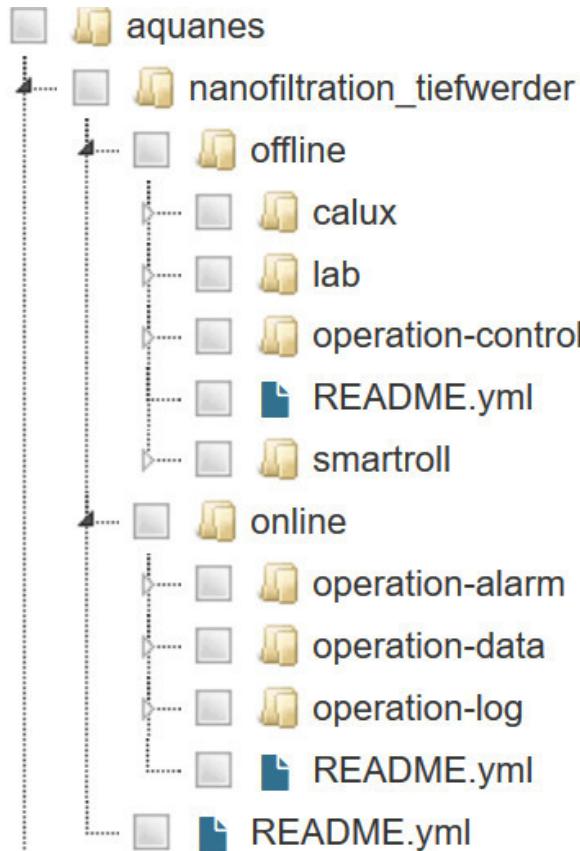
Langer Pfad zum Testprojekt





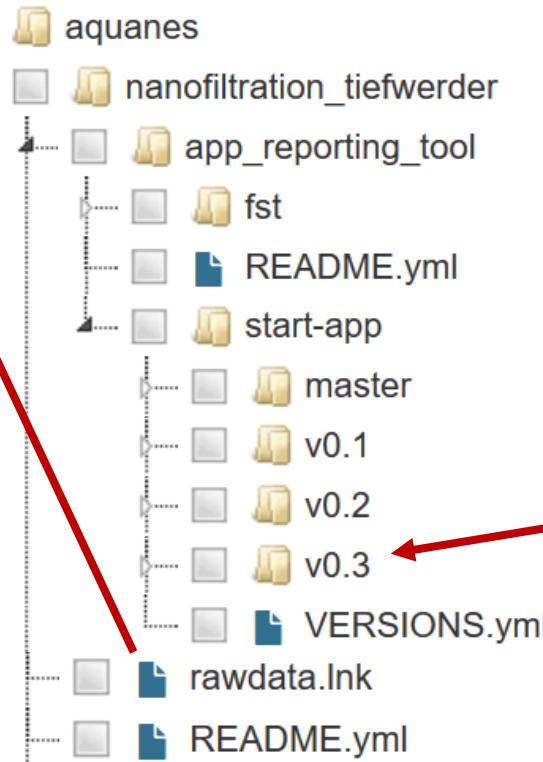
Rohdaten

\server\rawdata



Datenverarbeitung

\server\processing



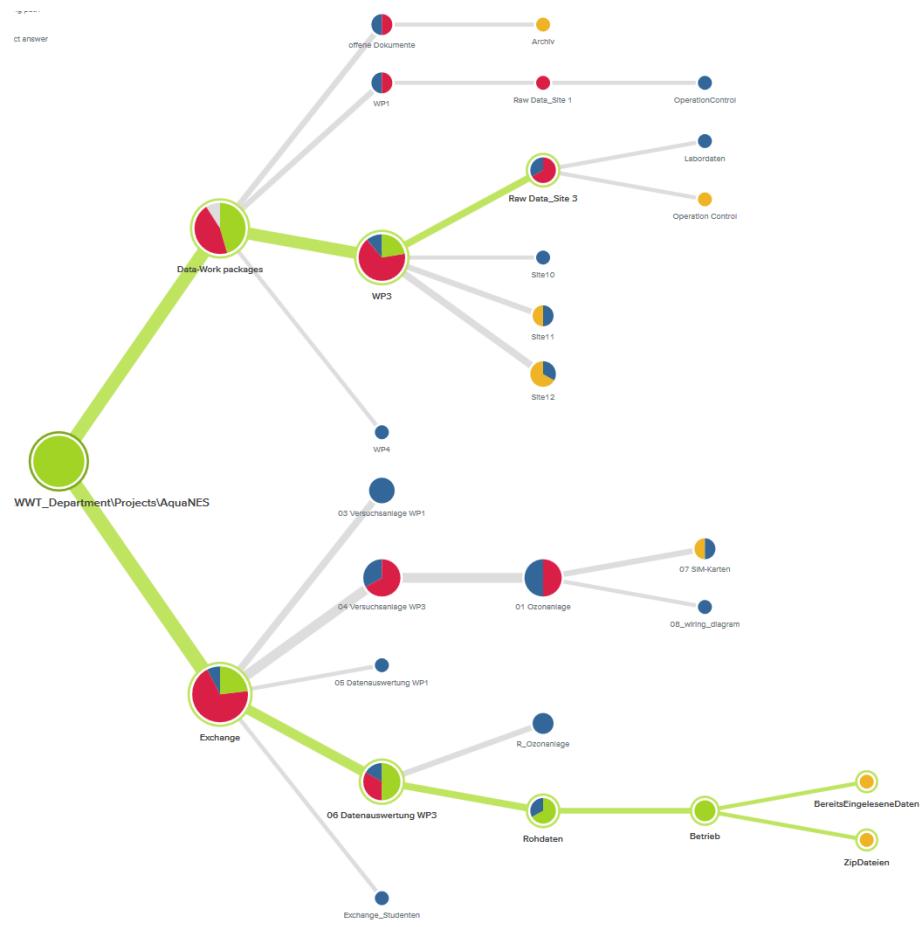
Ergebnisse

\server\projekte

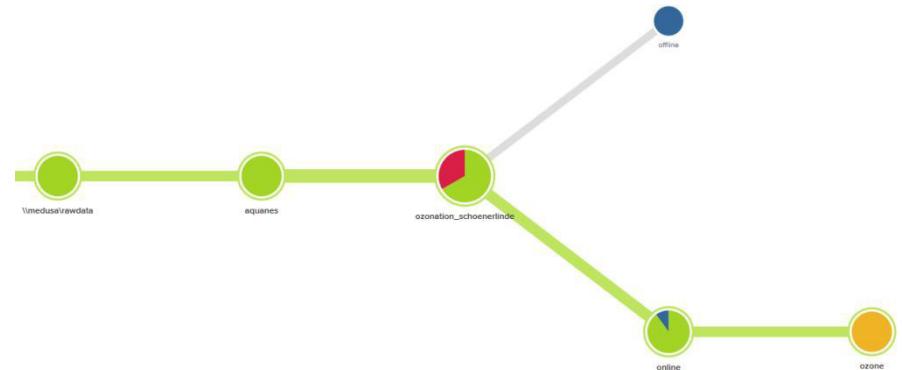
Link auf beste
verfügbare Version

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

Bisherige Ordnerstruktur



Neue Ordnerstruktur

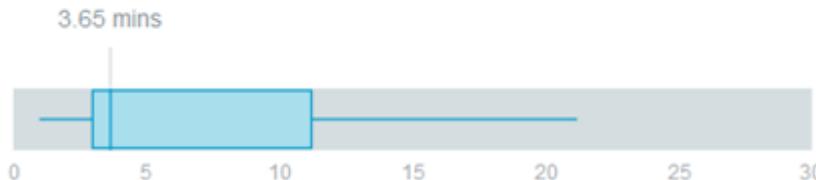


Wo finde ich die Rohdaten des Betriebes der Ozonanlage in Schönerlinde, die im Projekt AquaNes erhoben wurden?

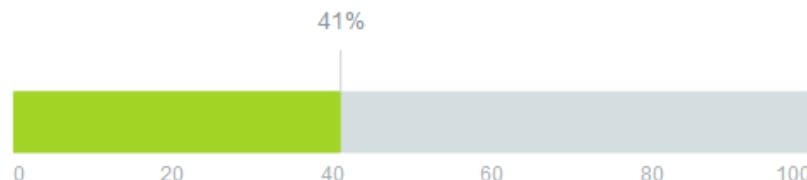
Bisherige Ordnerstruktur

n = 9

Time taken ?



Success ?



Directness ?



Neue Ordnerstruktur

n = 9

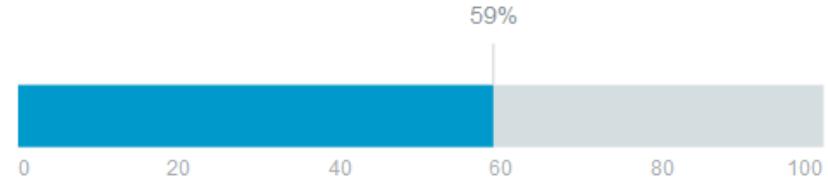
Time taken ?



Success ?



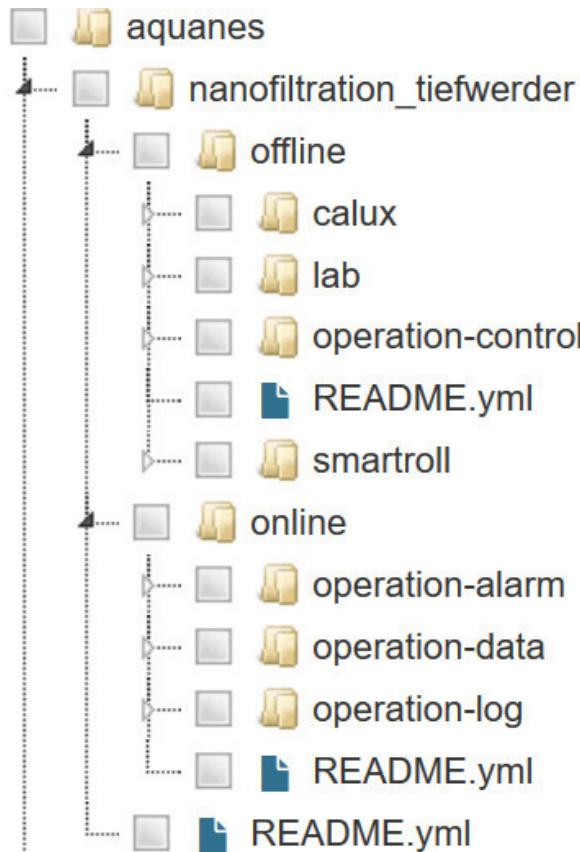
Directness ?





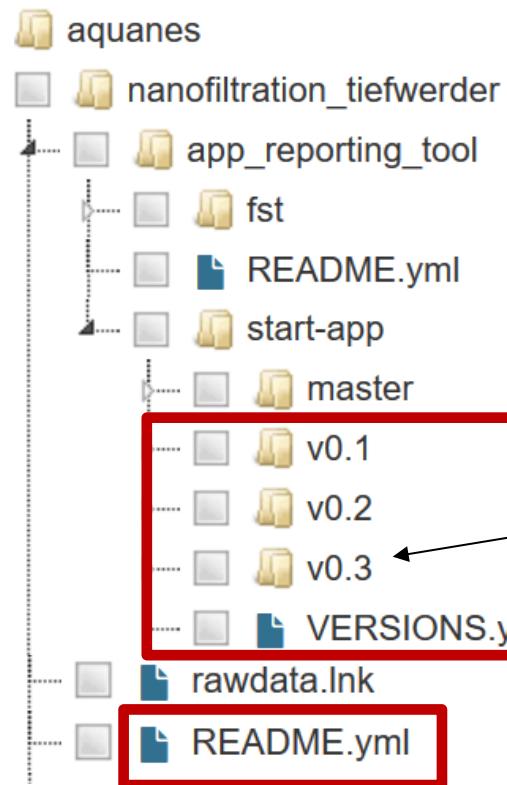
Rohdaten

\server\rawdata



Datenverarbeitung

\server\processing



Ergebnisse

\server\projekte

Link auf beste
verfügbare Version

Metadaten im YAML-Format

The screenshot shows a Notepad++ window displaying a YAML configuration file named README.yml. The file contains metadata for a pilot plant at Tiefwerder. The code is color-coded, with blue for keys and red for URLs. The file structure includes sections for rawdata, app_reporting_tool, and rawdata-import.

```
1 ---  
2 README.yml:  
3   created_by: "Michael Rustler"  
4   last_modified: 2019-03-06  
5   last_modified_by: "Michael Rustler"  
6  
7 app_reporting_tool:  
8   description: "App for nanofiltration pilot plant Tiefwerder"  
9   documentated_at: "https://github.com/KWB-R/aquanes.report"  
10  documented_by: "Michael Rustler"  
11  
12 rawdata:  
13   description: "link to \"read-only\" directory with rawdata for pilot plant  
14   Tiefwerder"  
15   documented_at: \\medusa\\rawdata\\aquanes\\nanofiltration_tiefwerder  
16   documented_by: "Michael Rustler"  
17  
18 rawdata-import:  
19   description: "folder with identical subfolder structure as \"rawdata\"  
20   containing only the latest files, which should be moved to from time to  
21   time by the \"data-manager\" into the \"read-only\" rawdata directory.\n"  
22   documented_at: \rawdata  
23   documented_by: "Michael Rustler"
```

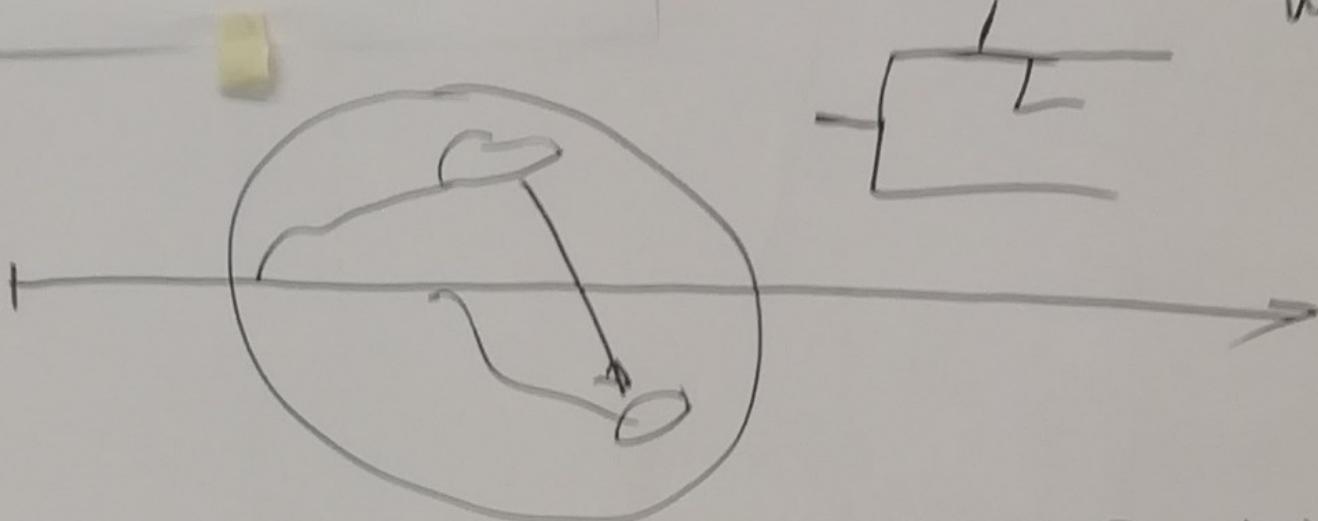
YAML Ain't Markup Language

length : 819 lines : 21 Ln : 17 Col : 1 Sel : 0 | 0 Windows (CR LF) UTF-8 INS



Lessons-learned Workshop

März 2019



Anschlussförderung
aktuellen Anlass
nutzen

wenig Speicher
„anbieten“

Bottom-up hat sich
bewährt

Lessons-learned

Datenstruktur: einfach, wenn neu, schwierig wenn laufend

Versionierung:

Mit Versionsverwaltung sind die Grundoperationen einfach, wenn gelernt (Subversion, Git)

Manuelle Versionierung erfordert Disziplin

Schwierig: Kollaboration mit externen Partnern (wenn nicht hinreichend „geschult“)

Bedarf: Datenmanagementplan, Softwaremanagementplan

Tool 1: Pfad-, Ordner- und Namenscheck

Analyse Paths

Table Sankey Explore Treemap Files in depth Stats Wordcloud

Load saved paths from

20190715_WWT_Department

Paths are relative to: '//medusa/projekte\$/WWT_Department'.

Search:

	toplevel	folder	filename	extension	type	size	modified	depth
	<input type="button" value="X"/>	AquaNES <input type="button" value="X"/>	All	A	<input type="button" value="X"/>	.	/	.
40929	Projects	AquaNES	Administration		directory	0	2017-09-04	3
40930	Projects	AquaNES	Closure & Evaluation		directory	0	2012-03-06	3
40931	Projects	AquaNES	Communication		directory	0	2019-03-19	3
40932	Projects	AquaNES	Data-Work packages		directory	0	2019-06-05	3
40933	Projects	AquaNES	Exchange		directory	0	2019-07-10	3
40934	Projects	AquaNES	Reports		directory	0	2019-07-04	3
40937	Projects	AquaNES/Communication	3D-Animation NF		directory	0	2019-05-07	4
40938	Projects	AquaNES/Communication	Docs		directory	0	2019-06-20	4
40939	Projects	AquaNES/Communication	Logos		directory	0	2018-08-09	4
40940	Projects	AquaNES/Communication	Pictures		directory	0	2018-10-10	4

Showing 1 to 10 of 1,236 entries (filtered from 97,357 total entries)

Previous

1

2

3

4

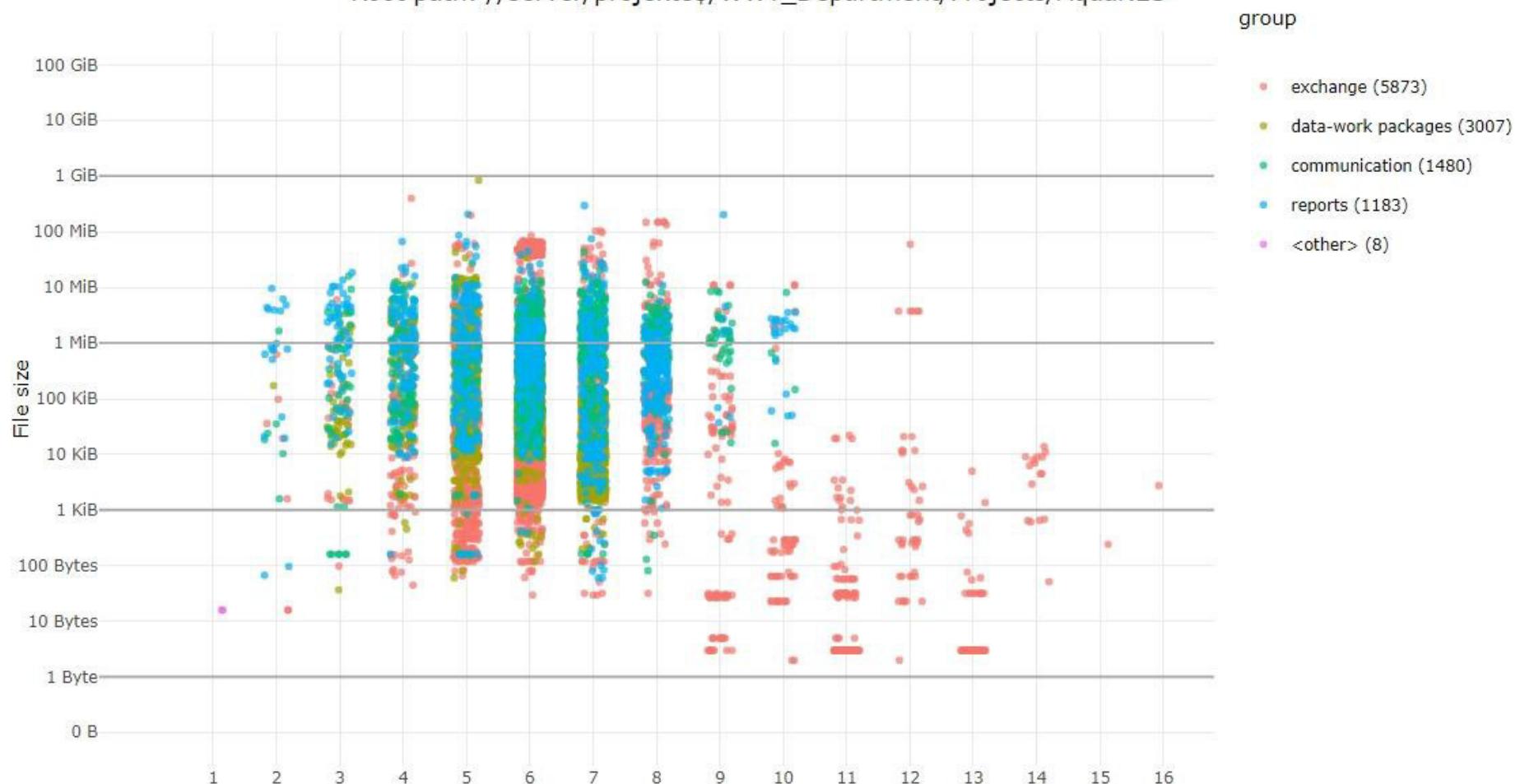
5

...

124

Next

Root path: //server/projekte\$/WWT_Department/Projects/AquaNES



Tool 2:

Der Wissensspeicher

Status Report

2019-07-24 07:30:02



<https://kwb-r.github.io/status/>

Repository	License	License_Badge	Tests_Coverage.io	Build_Windows	Build_Linux	Released_on_CRAN	Citation_DigitalObjectIdentifier	Doc_Rel
algoliar	MIT					CRAN	not published	
aquanes.report	MIT					CRAN	not published	DOI 10.5281/zenodo.1244103
fhpredict	MIT					CRAN	not published	
kwb.base	MIT					CRAN	not published	
kwb.code	MIT					CRAN	not published	
kwb.datetime	MIT					CRAN	not published	
kwb.db	MIT					CRAN	not published	
kwb.default	MIT					CRAN	not published	
kwb.demeau	MIT					CRAN	not published	
kwb.dwa.m150	MIT					CRAN	not published	
kwb.dwd	MIT					CRAN	not published	
kwb.en13508.2	MIT					CRAN	not published	
kwb.endnote	MIT					CRAN	not published	
kwb.event	MIT					CRAN	not published	
kwb.fakin	MIT					CRAN	not published	DOI 10.5281/zenodo.1309312
kwb.file	MIT					CRAN	not published	
kwb.geosalz	MIT					CRAN	not published	DOI 10.5281/zenodo.2563870
kwb.hantush	MIT					CRAN	0.2.1	DOI 10.5281/zenodo.61613
kwb.lca	MIT					CRAN	not published	
kwb.logger	MIT					CRAN	not published	DOI 10.5281/zenodo.1289425
kwb.monitoring	MIT					CRAN	not published	

Status Report

2019-07-24 07:30:02



<https://kwb-r.github.io/status/>

Repository	License	License_Badge	Tests_Coverage.io	Build_Windows	Build_Linux	Released_on_CRAN	Citation_DigitalObjectIdentifier	Doc_Rel
algoliar	MIT					CRAN	not published	
aquanes.report	MIT					CRAN	not published	
fhpredict	MIT					CRAN	not published	
kwb.base	MIT					CRAN	not published	
kwb.code	MIT					CRAN	not published	
kwb.datetime	MIT					CRAN	not published	
kwb.db	MIT					CRAN	not published	
kwb.default	MIT					CRAN	not published	
kwb.demeau	MIT					CRAN	not published	
kwb.dwa.m150	MIT					CRAN	not published	
kwb.dwd	MIT					CRAN	not published	
kwb.en13508.2	MIT					CRAN	not published	
kwb.endnote	MIT					CRAN	not published	
kwb.event	MIT					CRAN	not published	
kwb.fakin	MIT					CRAN	not published	
kwb.file	MIT					CRAN	not published	
kwb.geosalz	MIT					CRAN	not published	
kwb.hantush	MIT					CRAN	0.2.1	
kwb.lca	MIT					CRAN	not published	
kwb.logger	MIT					CRAN	not published	
kwb.monitoring	MIT					CRAN	not published	

Search...

Collects, aggregates and visualises operational analytical data from water suppliers (including a standardised reporting document)

Demo of reporting tool for Haridwar site

If you are interested in the capabilities of the R reporting tool, just make sure your computer satisfies the [software requirements](#) and subsequently execute the following lines of code below in R/RStudio.

```
if (!require("remotes")) {  
  install.packages("remotes", repos = "https://cloud.r-project.org")  
}  
  
remotes::install_github("kwb-r/aquanes.report",  
                        dependencies = TRUE)  
  
aquanes.report::run_app()
```

This will launch the R reporting tool in your default webbrowser. The demo is based on real operational and analytical data for the AQUANES site Haridwar for the time period from 2017-03-24 until 2017-07-25.

Browse source code at
<https://github.com/kwb-r/aquanes.report>

Report a bug at
<https://github.com/kwb-r/aquanes.report/issues>

License

[Full license](#)

[MIT + file LICENSE](#)

Developers

[Michael Rustler](#)

Author, maintainer 

 [AQUANES](#)

Funder

[All authors...](#)

Dev status









DOI [10.5281/zenodo.1244103](https://doi.org/10.5281/zenodo.1244103)

Status Report

2019-07-24 07:30:02



<https://kwb-r.github.io/status/>

Repository	License	License_Badge	Tests_Coverage.io	Build_Windows	Build_Linux	Released_on_CRAN	Citation_DigitalObjectIdentifier	Doc_Rel	
algoliar	MIT					CRAN	not published	X	
aquunes.report	MIT					CRAN	not published		X
fhpredict	MIT					CRAN	not published	X	
kwb.base	MIT					CRAN	not published	X	
kwb.code	MIT					CRAN	not published	X	
kwb.datetime	MIT					CRAN	not published	X	
kwb.db	MIT					CRAN	not published	X	
kwb.default	MIT					CRAN	not published	X	
kwb.demeau	MIT					CRAN	not published	X	
kwb.dwa.m150	MIT					CRAN	not published	X	
kwb.dwd	MIT					CRAN	not published	X	
kwb.en13508.2	MIT					CRAN	not published	X	
kwb.endnote	MIT					CRAN	not published	X	
kwb.event	MIT					CRAN	not published	X	
kwb.fakin	MIT					CRAN	not published		X
kwb.file	MIT					CRAN	not published	X	
kwb.geosalz	MIT					CRAN	not published		X
kwb.hantush	MIT					CRAN	0.2.1		X
kwb.lca	MIT					CRAN	not published	X	
kwb.logger	MIT					CRAN	not published		X
kwb.monitoring	MIT					CRAN	not published	X	

May 9, 2018

aquanes.report (v0.5.0)

Michael Rustler

Official release for AQUANES

- Haridwar (site 5): completed
- Berlin (sites 1 & 12): integrated with performance optimisation (but without analytics)
- Basel (site 6): integrated operational and analytical data for Wiese/Rhine sites (with new metadata for analytics)

Preview

aquanes.report-v.0.5.0.zip

KWB-R-aquanes.report-70fb50f

◦ .Rbuildignore	143 Bytes
◦ .gitignore	500 Bytes
◦ .travis.yml	193 Bytes
◦ DESCRIPTION	1.4 kB
◦ LICENSE	72 Bytes
◦ LICENSE.md	1.1 kB
◦ NAMESPACE	2.6 kB
◦ NEWS.md	2.3 kB
◦ R	
▪ aggregate_export_fst.R	7.1 kB
▪ calculate_operational_parameters_berlin_s.R	2.6 kB
▪ calculate_operational_parameters_berlin_t.R	1.9 kB
▪ calculate_operational_parameters_haridwar.R	8.5 kB
▪ check_thresholds.R	2.3 kB
▪ create_monthly_selection.R	909 Bytes
▪ dygraph_add_limits.R	1.4 kB
▪ def monthly data from calenderweeks R	3.3 kB

Files (25.6 MB)

Name

Size

[KWB-R/aquanes.report-v.0.5.0.zip](#)

25.6 MB

Preview Download

md5:70cf266bad27635e90f16afa99413168

Beta

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0)

Search



Citations to this version

Software Open Access

120

views

4

downloads

[See more details...](#)

Available in

GitHub

Indexed in

OpenAIRE**Publication date:**

May 9, 2018

DOI:[DOI 10.5281/zenodo.1244103](#)**Keyword(s):**

R package

Grants:[European Commission:](#)

- AquaNES - Demonstrating synergies in combined natural and engineered processes for water treatment systems (689450)

Related identifiers:

Documented by:

[https://kwb-r.github.io/aquanes.report/](#)

Supplement to:

[https://github.com/KWB-R/aquanes.report/tree/v.0.5.0](#)**Communities:**[Kompetenzzentrum Wasser Berlin gGmbH](#)

May 9, 2018

aquanes.report (v0.5.0)

Michael Rustler

Official release for AQUANES

- Haridwar (site 5): completed
- Berlin (sites 1 & 12): integrated with performance optimisation (but without analytics)
- Basel (site 6): integrated operational and analytical data for Wiese/Rhine sites (with new metadata for analytics)

Preview

aquanes.report-v.0.5.0.zip

KWB-R-aquanes.report-70fb50f

◦ .Rbuildignore	143 Bytes
◦ .gitignore	500 Bytes
◦ .travis.yml	193 Bytes
◦ DESCRIPTION	1.4 kB
◦ LICENSE	72 Bytes
◦ LICENSE.md	1.1 kB
◦ NAMESPACE	2.6 kB
◦ NEWS.md	2.3 kB
◦ R	
▪ aggregate_export_fst.R	7.1 kB
▪ calculate_operational_parameters_berlin_s.R	2.6 kB
▪ calculate_operational_parameters_berlin_t.R	1.9 kB
▪ calculate_operational_parameters_haridwar.R	8.5 kB
▪ check_thresholds.R	2.3 kB
▪ create_monthly_selection.R	909 Bytes
▪ dygraph_add_limits.R	1.4 kB
▪ def_monthly_data_from_calenderweeks.R	3.3 kB

Files (25.6 MB)

Name

Size

[KWB-R/aquanes.report-v.0.5.0.zip](#)

25.6 MB

Preview Download

md5:70cf266bad27635e90f16afa99413168

Beta

Citations 0

Show only: Literature (0) Dataset (0) Software (0) Unknown (0)

Search



Citations to this version

Software Open Access

120

views

4

downloads

[See more details...](#)

Available in

GitHub

Indexed in

OpenAIRE**Publication date:**

May 9, 2018

DOI:[DOI 10.5281/zenodo.1244103](#)**Keyword(s):**

R package

Grants:European Commission:

- AquaNES - Demonstrating synergies in combined natural and engineered processes for water treatment systems (689450)

Related identifiers:

Documented by:

[https://kwb-r.github.io/aquanes.report/](#)

Supplement to:

[https://github.com/KWB-R/aquanes.report/tree/v0.5.0](#)**Communities:**[Kompetenzzentrum Wasser Berlin gGmbH](#)



Search



Upload

Communities

 michael.rustler@kompetenz-wasser.de

Kompetenzzentrum Wasser Berlin gGmbH

Recent uploads

Search Kompetenzzentrum Wasser Berlin gGmbH



 New upload

February 13, 2019 (v0.1.0) Software Open Access

kwb.geosalz (v0.1.0): R Package for Documenting Workflow Used in Project "geosalz"

[View](#)

 Michael Rustler,  Hauke Sonnenberg;

R Package for Documenting Workflow Used in Project "geosalz"

Uploaded on February 13, 2019

December 13, 2018 (v0.1.0) Software Open Access

[View](#)

kwb.resilience (v0.1.0): R Package for the Quantification of Technical Resilience

 Andreas Matzinger,  Michael Rustler,  Hauke Sonnenberg;

Documentation website: <https://kwb-r.github.io/kwb.resilience>

Uploaded on December 13, 2018

November 19, 2018 (v0.4.0) Software Open Access

[View](#)

kwb.utils (v0.4.0)

 Hauke Sonnenberg,  Michael Rustler,

Community

**KOMPETENZZENTRUM
WasserBerlin**

Kompetenzzentrum Wasser Berlin gGmbH

Publications by Researchers Working At
[Kompetenzzentrum Wasser Berlin gGmbH]
(<https://kompetenz-wasser.de>)

Curated by:

mrustl

Curation policy:

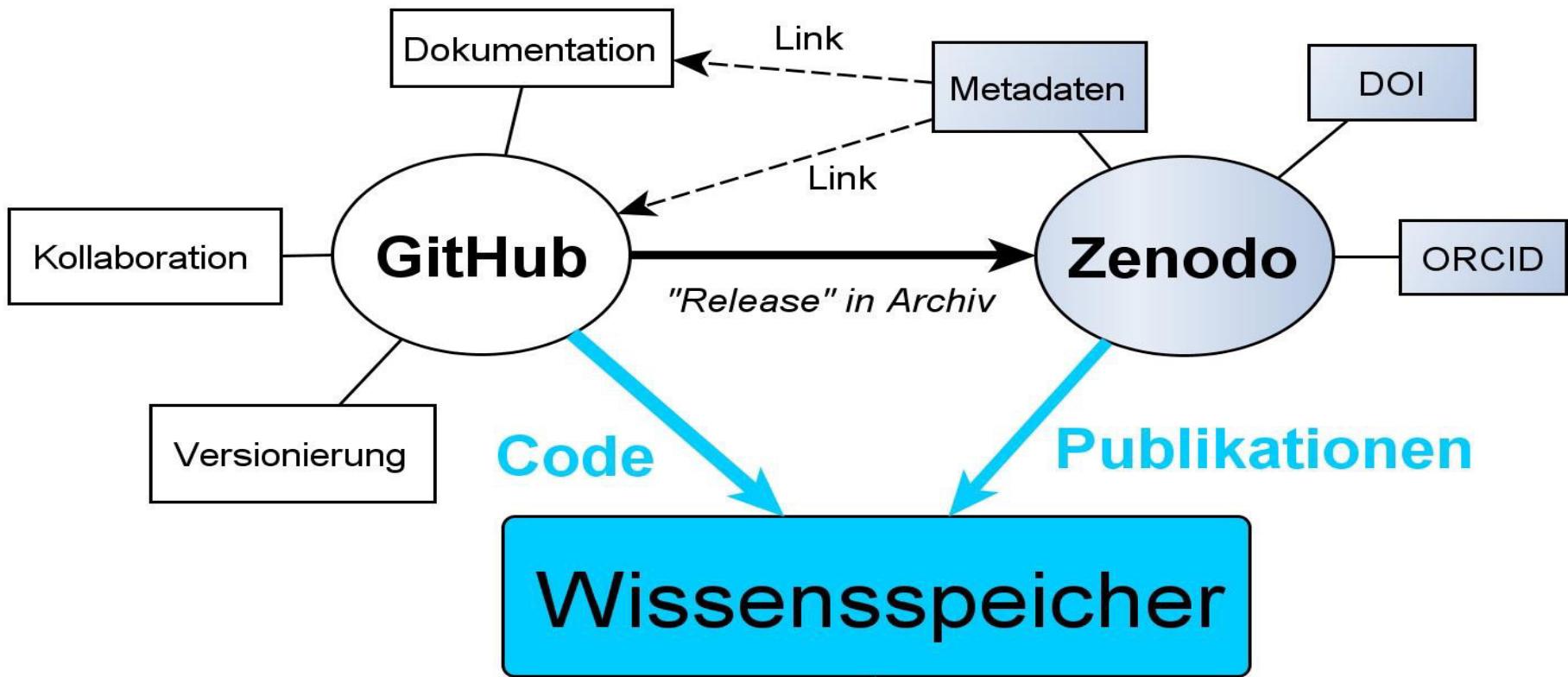
Not specified

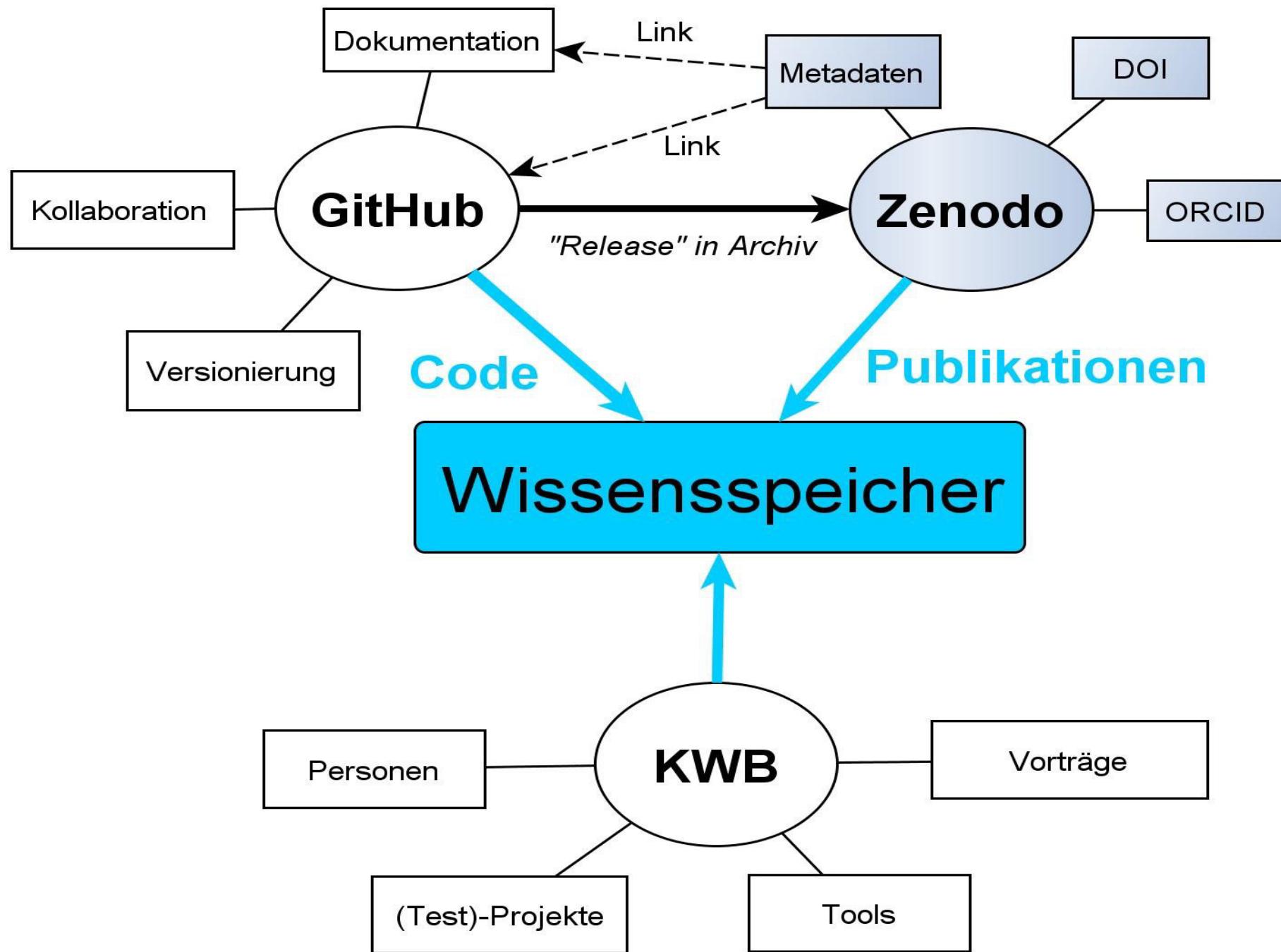
Created:

February 23, 2019

Harvesting API:

OAI-PMH Interface





FAKIN (Research Data Management At Small Institutes)

Development of standardised processes, tools and methods for
the reproducibility of research results

photo: Jürgen Stamp, digitalbevaring.dk

Keyfacts



Duration

May 2017 - July 2019



Volume

157.665 €



Funder

BMBF

Hintergrund: Wissensspeicher

FAKIN Latest Tools Projects Apps Code People Publications Tags Talks

Knowledge Repo: An Innovative Way For Sharing Knowledge At An Institutional Level

Michael Rustler, Hauke Sonnenberg

['A' OpenScienceRadio Interview](#)

[Etherpad: Session 10 "Knowledge Repo"](#)

[Etherpad: Sessions Metapad](#)

[Link to Barcamp](#)

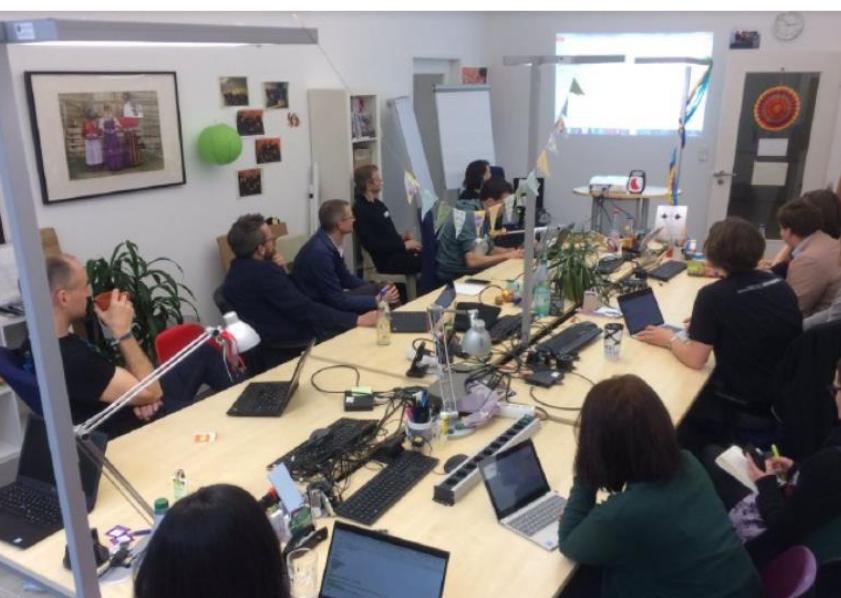
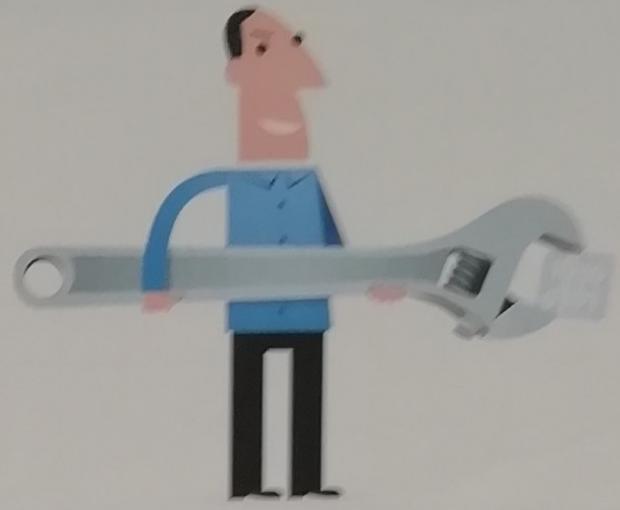


Image credit: Sarah-Isabella Behrens

Date	Feb 16, 2019 2:45 PM
Event	Barcamp Open Science
Location	Wikimedia Deutschland, Tempelhofer Ufer 23-24

Toolsammlung

Card Sorting



Datenschule Zenodo

-ime Survey

Subversion DMP Online

Git Hub
Git Lab

RDM Organizer

treejack

Software Management Plan

YAML

Powershell

NotePad++

DataCamp

Hugo

Wissensmanager

Linksammlung Tools

Erstellung von Forschungsdatenmanagementplänen

- DMP Online: <https://dmponline.dcc.ac.uk>
- RDMOrganizer: <https://rdmorganiser.github.io>
- Software Management Plan: <https://doi.org/10.5281/zenodo.1422656>

Evaluation der „Datenkompetenz“ im Institut

- Data Literacy Matrix der „Datenschule“: <https://datenschule.de/workshops>

Kollaborative Programmierplattformen

- GitHub: <https://github.com>
- Gitlab: <https://gitlab.com>

Metadatenformat

- YAML-Format: <https://de.wikipedia.org/wiki/YAML>
- YAML-Formatcheck: <http://www.yamlint.com>

Online-Umfrage

- LimeSurvey: <https://limesurvey.org>

Programmieren lernen (z. B. R, Python)

- Datacamp (Online-Lernplattform): <https://datacamp.com>

Repositorien (Archivierung, Nachnutzung)

- re3data.org (Forschungsrepositorien-Suchmaschine): <https://re3data.org>
- Zenodo (allgemeines Repozitorium): <https://zenodo.org>

Testen von Ordnerstrukturen

- Treejack: <https://www.optimalworkshop.com/treejack>
- Card sorting: <https://www.optimalworkshop.com/optimalsort>

Texteditor

- Notepad++ („versteht“ auch YAML Format): <https://notepad-plus-plus.org>

Versionsverwaltungssoftware

- Subversion: <https://subversion.apache.org>
- Git: <https://git-scm.com>

Wissensspeicher Generierung

- Hugo (statischer Webseitengenerator): <https://gohugo.io>
- hugo-academic (Open Source-Webseitenvorlage): <https://sourcethemes.com/academic>