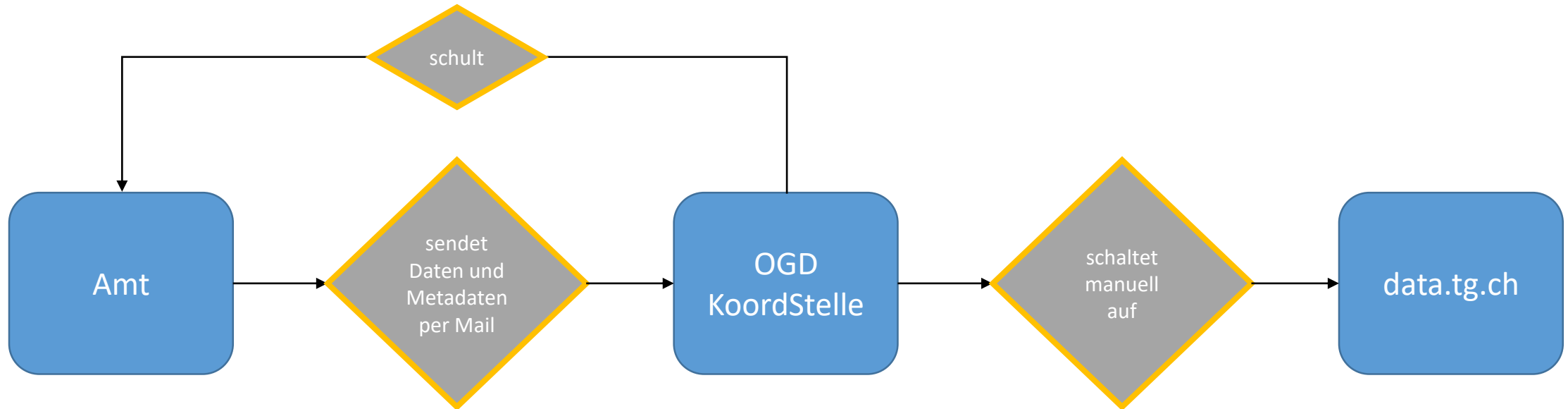


Wer braucht schon Python?

Automatisierte Workflows mit R und Gitlab CI/CD

Step 0: Ausgangszustand



Die Problemstellung

- Kleines Team bei OGD-Thurgau
- Mit der OGD-Strategie wird auch die Zahl der Datensätze steigen
- Manuelle Aufschaltung ist zeitaufwändig und fehleranfällig

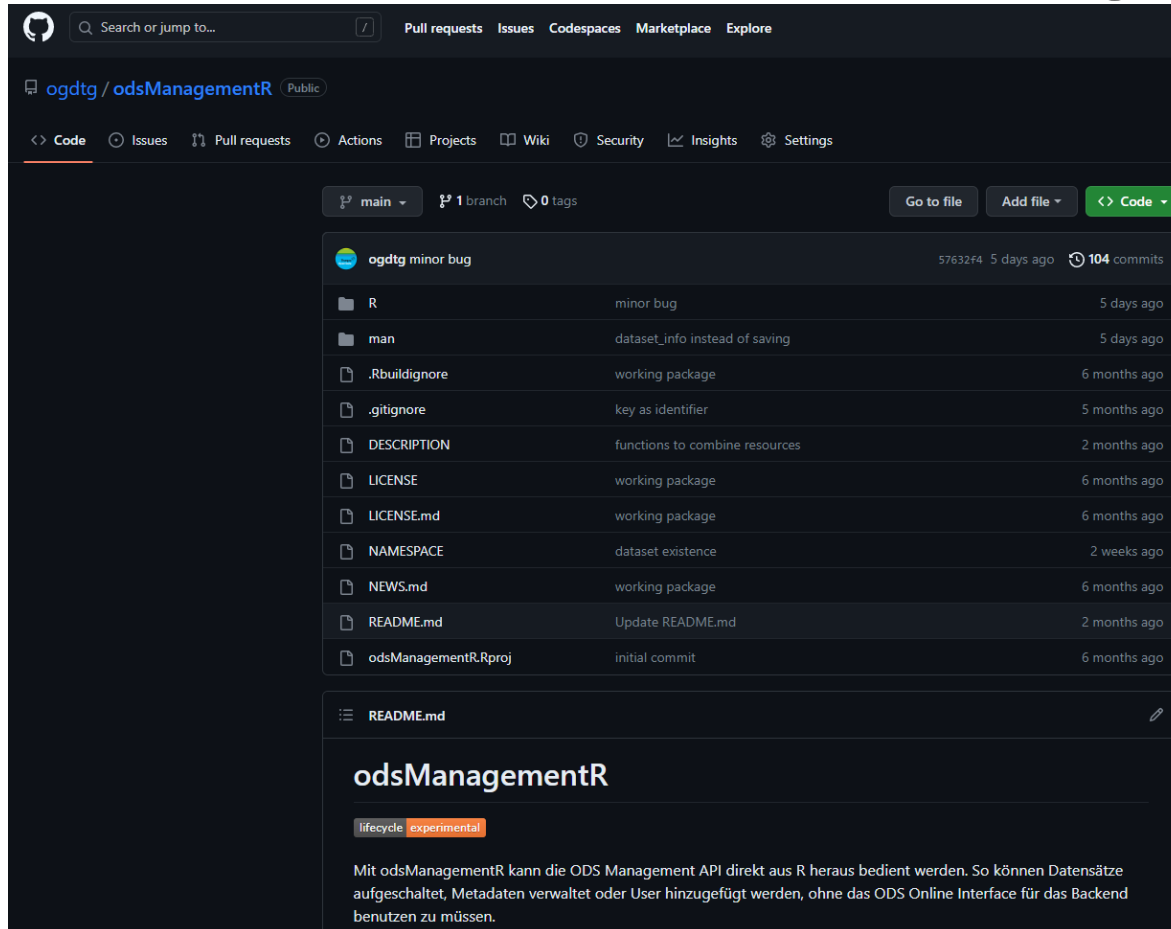
Können wir diesen Prozess automatisieren?

Step 1: API zu data.tg.ch

- Schnittstelle wird vom Betreiber opendatasoft zur Verfügung gestellt
- Aufschalten von Daten und Metadaten per API möglich

Entwicklung des odsManagementR Package, um API direkt aus R heraus anzusteuern

Step 1: API zu data.tg.ch



ogdtg / odsManagementR Public

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags Go to file Add file Code

ogdtg minor bug 57632f4 5 days ago 104 commits

R	minor bug	5 days ago
man	dataset_info instead of saving	5 days ago
.Rbuildignore	working package	6 months ago
.gitignore	key as identifier	5 months ago
DESCRIPTION	functions to combine resources	2 months ago
LICENSE	working package	6 months ago
LICENSE.md	working package	6 months ago
NAMESPACE	dataset existence	2 weeks ago
NEWS.md	working package	6 months ago
README.md	Update README.md	2 months ago
odsManagementR.Rproj	initial commit	6 months ago

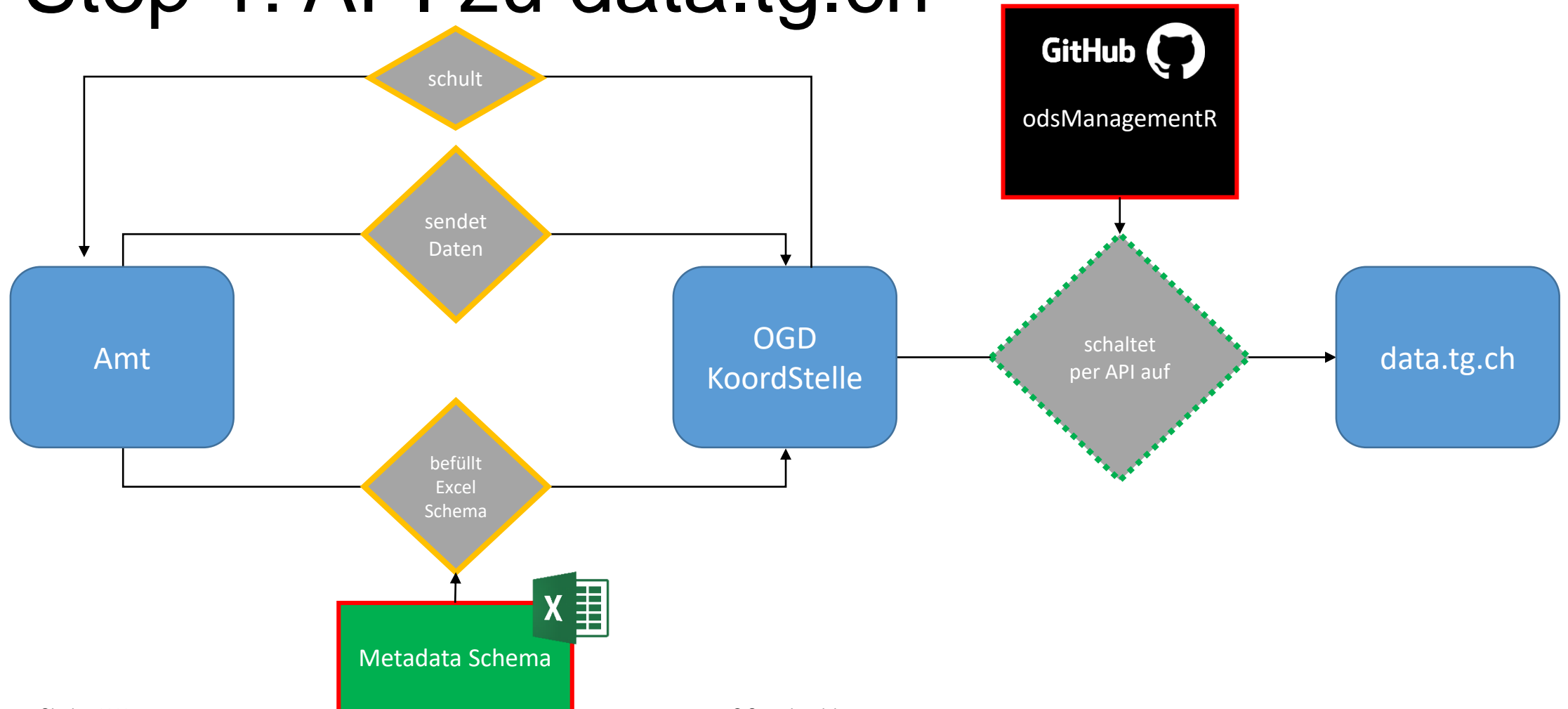
README.md

odsManagementR

lifecycle experimental

Mit odsManagementR kann die ODS Management API direkt aus R heraus bedient werden. So können Datensätze aufgeschaltet, Metadaten verwaltet oder User hinzugefügt werden, ohne das ODS Online Interface für das Backend benutzen zu müssen.

Step 1: API zu data.tg.ch



Step 1: API zu data.tg.ch

1	Metadaten	Eintrag
2	Departement	DEK
3	Amt	Amt für Berufsbildung und Berufsberatung
4	Abkürzung	ABB
5	Kennung	dek-abb
6		Abschlüsse Berufliche Grundbildung seit 2017
7		Abschlüsse Berufliche Grundbildung seit 2017
8		Abschlüsse Berufliche Grundbildung seit 2017
9		Abschlüsse Berufliche Grundbildung seit 2017
10		Abschlüsse Berufliche Grundbildung seit 2017
11		Abschlüsse Berufliche Grundbildung seit 2017
12		Abschlüsse Berufliche Grundbildung seit 2017
13		Abschlüsse Berufliche Grundbildung seit 2017
14		Abschlüsse Berufliche Grundbildung seit 2017
15		Abschlüsse Berufliche Grundbildung seit 2017

	A	B	C	D	E	F	G	H	I
1	berufsnr	isced	berufsbezeichnung	efz_eba	jahr	bildungsverh_m	bildungsverh_w	ausbildungsverh_vert	
2	54602	Architektur und Städteplanung	Architekturmodellbauer	EBA	2017	0,1,1,0,0,1,1,1,0,0			
3	54603	Architektur und Städteplanung	Architekturmodellbauer EFZ	EFZ	2017	0,1,0,0,0,0,0,0,0,1			
4	64008	Architektur und Städteplanung	Zeichner EFZ Fachrichtung Architektur	EFZ	2017	65,31,23,9,1,3,24,12,21,11			
5	64009	Architektur und Städteplanung	Zeichner EFZ Fachrichtung Ingenieurbau	EFZ	2017	42,12,10,4,1,0,11,4,11,4			
6	64010	Architektur und Städteplanung	Zeichner EFZ Fachrichtung Innenarchitektur	EFZ	2017	3,3,0,3,0,3,0,0,0,3,0,3,0			
7	64104	Architektur und Städteplanung	Geomatiker EFZ amtliche Vermessung	EFZ	2017	17,6,8,0,0,0,0,0,0,5,2			
8	64104	Architektur und Städteplanung	Geomatiker EFZ Geoinformatik	EFZ	2017	2,1,0,0,0,0,0,0,0,1,0			
9	33310	Audiovisuelle Techniken und Medienproduktion	Flexodrucker EFZ Letzter Lehrbeginn 2020	EFZ	2017	2,0			
10	34126	Audiovisuelle Techniken und Medienproduktion	Drucktechnologie EFZ Siebdruck	EFZ	2017	1,0,1,0,0,0,1,0			
11	34127	Audiovisuelle Techniken und Medienproduktion	Drucktechnologie EFZ Reprografie	EFZ	2017	0,1,1,0,0,0,1,0			
12	34128	Audiovisuelle Techniken und Medienproduktion	Drucktechnologie EFZ Bogendruck	EFZ	2017	11,3,0,0,0,0,0			
13	34129	Audiovisuelle Techniken und Medienproduktion	Drucktechnologie EFZ Rollendruck	EFZ	2017	0,0,1,0,0,0,0,0			
14	34708	Audiovisuelle Techniken und Medienproduktion	Polygraf EFZ Mediengestaltung	EFZ	2017	0,0,0,2,0,0,0,2			
15	34709	Audiovisuelle Techniken und Medienproduktion	Polygraf EFZ Medienproduktion	EFZ	2017	0,0,5,5,0,0,5,5			
16	34710	Audiovisuelle Techniken und Medienproduktion	Polygraf EFZ Printmedien Letzter Lehrbeginn 2021	EFZ	2017	0,0,0,0,0,0,0,0			
17	35211	Audiovisuelle Techniken und Medienproduktion	Fotofachmann EFZ Fotografie Letzter Lehrbeginn 2021	EFZ	2017	0,0,0,0,0,0,0,0			
18	35312	Audiovisuelle Techniken und Medienproduktion	Printmedienverarbeiter EFZ Bindetechnologie letzter Lehr	EFZ	2017	0,0,0,0,0,0,0,0			
19	35314	Audiovisuelle Techniken und Medienproduktion	Printmedienverarbeiter EFZ Versandtechnologie letzter Lehr	EFZ	2017	0,0,0,0,0,0,0,0			
20	35315	Audiovisuelle Techniken und Medienproduktion	Printmedienverarbeiter EFZ Druckausrüstung letzter Lehr	EFZ	2017	0,0,0,0,0,0,0,0			
21	35316	Audiovisuelle Techniken und Medienproduktion	Printmedienpraktiker EBA Kartonage und Verpackung	EBA	2017	0,0,0,0,0,0,0,0			
22	35316	Audiovisuelle Techniken und Medienproduktion	Printmedienpraktiker EBA Druckweiterverarbeitung	EBA	2017	0,0,0,0,0,0,0,0			
23	47117	Audiovisuelle Techniken und Medienproduktion	Interactive Media Designer EFZ	EFZ	2017	0,0,0,0,0,0,0,0,0			
24	47121	Audiovisuelle Techniken und Medienproduktion	Mediamatiker EFZ Letzter Lehrbeginn 2018	EFZ	2017	88			

```
3 # Metadaten hochladen
4 dataset_uid <- add_metadata_from_scheme(
5   filepath = "path/to/metadata.xlsx"
6 )
7
8 # Daten hinzufügen
9 add_data_to_dataset(dataset_uid = dataset_uid,
10  schema = "path/to/metadata.xlsx",
11  ogd_file = "path/to/ogd_data.csv",
12  resource_title = "Grundbildung seit 2017")
13
```

Wie können wir den Aufwand weiter verringern?

Abschlüsse Berufliche Grundbildung seit 2017 [Bearbeiten](#)

[Information](#) [Tabelle](#) [Exporte](#) [API](#)

Der Datensatz enthält die Totale Anzahl der Lehrverträge, sowie die neu registrierten Lehrverträge des entsprechenden Jahres.

Aufgeführt sind auch die Abschlüsse der Qualifikationsverfahren (Lehrabschlussprüfungen) nach Ausbildungsfeldern. Die Zahlen beinhalten Ausbildungsverhältnisse, Repetenten, durchgeführte Qualifikationsverfahren, sowie ausgestellte Fähigkeits- und Attestausweise. Basis für die Daten bilden die Aufbereitung der Jahresstatistik für das Bundesamt für Statistik.

Die Zahlen der Lehrvertragsauflösungen sind hier nicht veröffentlicht.

Zu kleine Werte für Repetenten, Prüfungen und ausgestellte Atteste können aus datenschutzrechtlichen Gründen nicht veröffentlicht werden.

Datenquelle: Amt für Berufsbildung und Berufsberatung

Datensatzkennung [dek-abb-1](#)

Downloads 8

Themen Bevölkerung, Bildung, Wissenschaft

Schlüsselwörter grundbildung, bildung, abschluss, abschlüsse, ausbildung, berufsbildung, beruf, handwerk, attest, prüfung, repetent

Lizenz CC BY 4.0

Sprache Deutsch

Zeitzone Europe/Zurich

Letzte Änderung 27. April 2023 10:00

Herausgeber Amt für Berufsbildung und Berufsberatung Kanton Thurgau

Referenz <https://abb.tg.ch/>

tags opendata.swiss

Gebiet [Thurgau](#)

berufsnr	isced	berufsbezeichnung	efz_eba	jahr	bildungsverh_m
1	54602	Architektur und Städteplanung	Architekturmodellbauer	EBA	2017 0
2	54603	Architektur und Städteplanung	Architekturmodellbauer EFZ	EFZ	2017 0
3	64010	Architektur und Städteplanung	Zeichner EFZ Fachrichtung Innena...	EFZ	2017 3
4	64104	Architektur und Städteplanung	Geomatiker EFZ Geoinformatik	EFZ	2017 2
5	33310	Audiovisuelle Techniken und Medi...	Flexodrucker EFZ Letzter Lehrbegi...	EFZ	2017 2
6	34708	Audiovisuelle Techniken und Medi...	Polygraf EFZ Mediengestaltung	EFZ	2017 0
7	34709	Audiovisuelle Techniken und Medi...	Polygraf EFZ Medienproduktion	EFZ	2017 0
8	35211	Audiovisuelle Techniken und Medi...	Fotofachmann EFZ Fotografie Letz...	EFZ	2017 4
9	47117	Audiovisuelle Techniken und Medi...	Interactive Media Designer EFZ	EFZ	2017 0
10	90305	Audiovisuelle Techniken und Medi...	Grafiker EFZ	EFZ	2017 1
11	39204	Baugewerbe, Hoch- und Tiefbau	Steinmetz EFZ Letzter Lehrbeginn...	EFZ	2017 0
12	47704	Baugewerbe, Hoch- und Tiefbau	Sanitärinstallateur EFZ Letzter Le...	EFZ	2017 82
13	47705	Baugewerbe, Hoch- und Tiefbau	Hautechnikpraktiker EBA Sanitär	EBA	2017 13
14	51204	Baugewerbe, Hoch- und Tiefbau	Ofenbauer EFZ	EFZ	2017 2
15	51417	Baugewerbe, Hoch- und Tiefbau	Grundbauer EFZ	EFZ	2017 4
16	51426	Baugewerbe, Hoch- und Tiefbau	Strassenbaupraktiker EBA	EBA	2017 6
17	51708	Baugewerbe, Hoch- und Tiefbau	Boden-Parkettleger EFZ Parkett L...	EFZ	2017 4
18	51912	Baugewerbe, Hoch- und Tiefbau	Polybauer EFZ Sonnenschutz-Sys...	EFZ	2017 5
19	51913	Baugewerbe, Hoch- und Tiefbau	Polybaupraktiker EBA Abdichten	EBA	2017 1
20	51913	Baugewerbe, Hoch- und Tiefbau	Polybaupraktiker EBA Dachdecken	EBA	2017 3
21	51913	Baugewerbe, Hoch- und Tiefbau	Polybaupraktiker EBA Gerüstbau	EBA	2017 0
22	51915	Baugewerbe, Hoch- und Tiefbau	Dachdecker EFZ	EFZ	2017 8

Idee

- Ämter sollen Daten selbst «hochladen»
- Aufschaltung soll automatisiert erfolgen, getriggert durch das Amt
- Ämter sollen nur Zugriff auf ihre Daten haben
- OGD-Koordinationsstelle soll Prozess überwachen und steuern können

Step 2: Automatisierung des Prozess

Was wir nicht haben:

- Zugang zu Servern, auf denen wir automatisiert Prozesse laufen lassen können
- Zugang zu FTP Servern, mit denen wir Datensätze harvesten können
- Python

Was wir haben:

- On-Premise Gitlab Instanz (gitlab.tg.ch)
- R und R Studio
- Git
- API Schnittstelle zu unserer OGD Plattform von opendatasoft

Step 2: Automatisierung des Prozess

Die Lösung: Gitlab CI/CD

- Tool zur Automatisierung von Tests und Deployment-Prozessen
- ermöglicht die Erstellung von Pipelines, die aus mehreren Schritten bestehen können
- bietet übersichtliche Benutzeroberfläche zur Überwachung der Pipelines
- ermöglicht die Einbindung von Docker Images wie z.B. **rocker**

Datenrepo und Coderepo

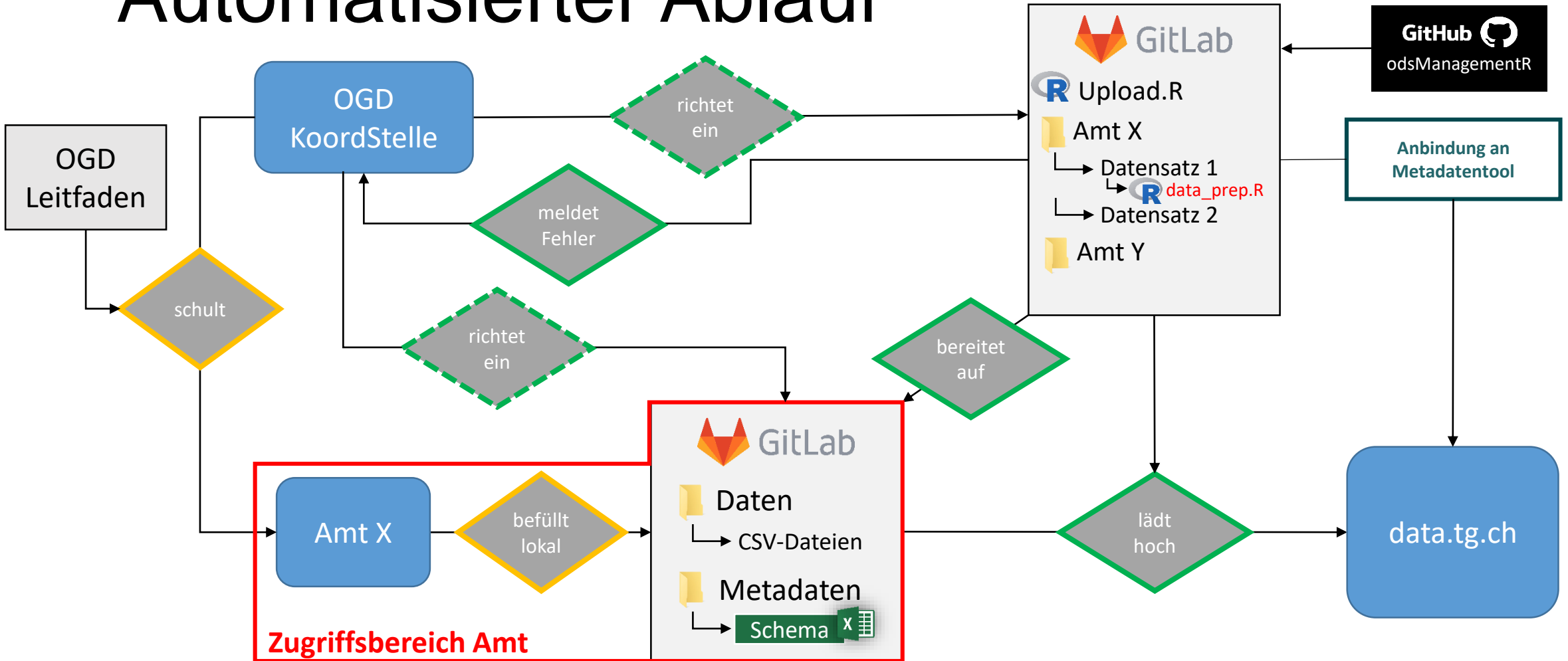
Datenrepo

- ein Datenrepo pro Amt
- Amt kann Daten und Metadaten ablegen
- enthält ein bash script zum vereinfachten pushen und pullen

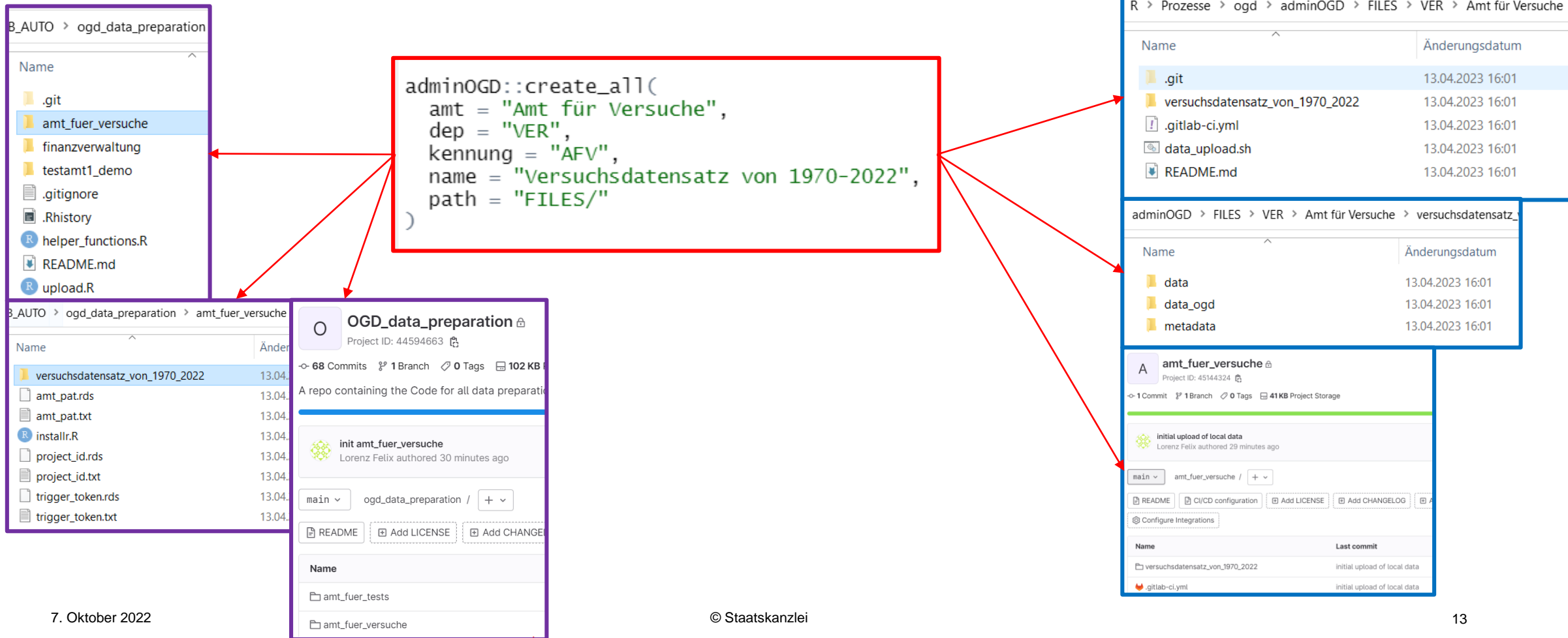
Coderepo

- enthält **Upload.R** File welches Prozess steuert
- enthält ein **data_prep.R** File pro Datensatz
- (noch) kein Zugriff durch das Amt

Automatisierter Ablauf



Repos einrichten mit adminOGD



```
adminOGD::create_all(  
  amt = "Amt für Versuche",  
  dep = "VER",  
  kennung = "AFV",  
  name = "Versuchsdatensatz von 1970-2022",  
  path = "FILES/"  
)
```

The diagram illustrates the setup of repositories using the `adminOGD` function. The central code block shows the configuration for creating a repository. Red arrows point from this code to four screenshots of the adminOGD interface, showing the resulting repository structure and content.

Repository Structure (Left Screenshot):

- ogd_data_preparation
 - amt_fuer_versuche
 - versuchsdatensatz_von_1970_2022
 - amt_pat.rds
 - amt_pat.txt
 - installr.R
 - project_id.rds
 - project_id.txt
 - trigger_token.rds
 - trigger_token.txt

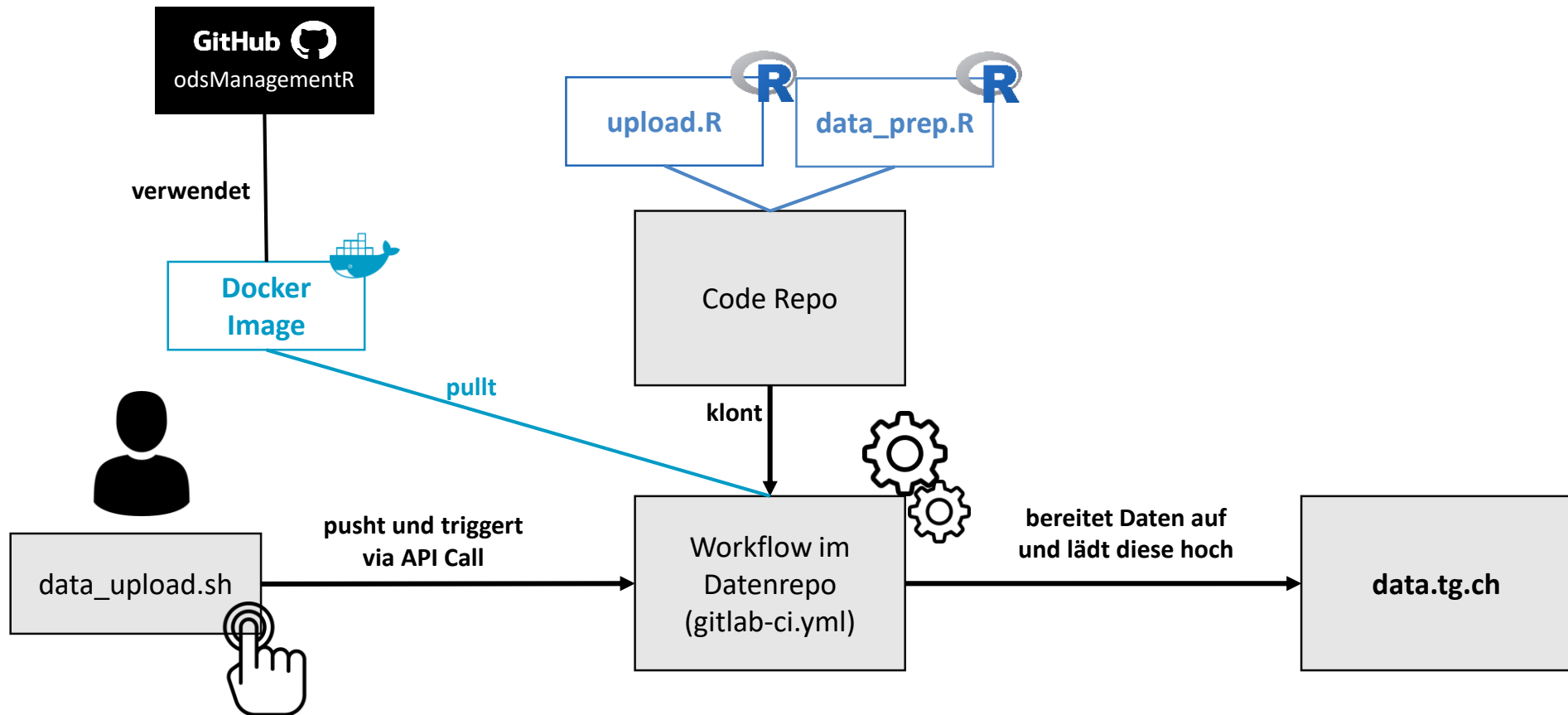
Repository Content (Right Screenshot):

- FILES > VER > Amt für Versuche
 - versuchsdatensatz_von_1970_2022
 - data
 - data_ogd
 - metadata

Repository Details (Bottom Screenshot):

- amt_fuer_versuche
 - initial upload of local data
 - main branch: amt_fuer_versuche /
 - Files: README, Add LICENSE, Add CHANGELOG
 - Commits: 1 Commit

Das Amt als Auslöser



Ergebnis

- manuelle Arbeit für OGD-Team und Fehleranfälligkeit minimiert
- kein wirklicher Mehraufwand für die Ämter
- Selbstverwaltete Pipelines, die nach Belieben angepasst werden können
- Möglichkeit für OGD-Team, Prozesse zu überwachen und ggf. einzugreifen
- Nutzer verwenden Git, ohne Git wirklich verstehen zu müssen

Learnings

- **Logging, Logging, Logging**
 - [lgr-Package](#) hat sich bewährt
- **Lokal und nicht remote testen**
 - Schreibt Skripte so, dass sie lokal getestet werden können
- **Flussdiagramme und generell Visualisierungen helfen dabei, Fehler in Prozessen zu ermitteln**
 - das [flow-Package](#) kann ausserdem euren Code visualisieren

Veranstaltungshinweis

OPEN DATA BEER Vol. 22:

WO: Bierlager, Brauhaus Frauenfeld

WANN: 14.06.2023, 18 Uhr

