

Verbreitung von Familiennamen (working title)

Projektskizze für #cdvnord

<https://codingdavinci.de>

Jörn Dinkla, joern@dinkla.net

18. September 2016

Projektidee

- Grundlage: (gestern Jesper Zedlitz)
Daten der Deutschen Arbeitsgemeinschaft
Genealogischer Verbände (DAGV)
- Ziel
Web-Seite für “Visualisierung”
(Zeit, Raum und Namensänderungen)

Die Daten

- Nachname, Zeitraum, Ort

ID	FAMILYNAME	BEGIN_DATE	END_DATE	SUBMITTER_ID	DENOMINATION	COUNTRY	REGION	POSTALCODE	PLACE_NAME	PLACE_UP
53	53	Schröder	1800	1900	1199	D	SH	23701	Roge	
54	54	Holling	1700	2000	1199	D	SH	24594	Tappendorf	http://gov
55	55	Barg	1800	1900	1199	D	SH	24594	Tappendorf	http://gov
56	56	Stabe	1800	2000	1199	D	SH	24802	Bokel	http://gov
57	57	Grage	1700	1900	1199	D	SH	24802	Bokel	http://gov
58	58	Aal	1666	9999	1318	D	He	64404	Bickenbach	http://gov
59	59	Aalst, van	1898	1930	798	NL				
60	60	Aart	1727	1799	528	D	Ni	49832	Freren, Krs. Emsland	
61	61	Abbenthaler	1716	1743	331	lu	D	27793	Wildeshausen	http://gov
62	62	Abbenthalern	1828	1912	1435	D	Ni	30159	Hannover	http://gov
63	63	Abeck	1860	1887	409	rk	SL	66740	Neuforweiler	http://gov
64	64	Abehus	1783	9999	331	D	Ni	27793	Wildeshausen	http://gov
65	65	Abel	1702	9999	682	D	Ni	30159	Hannover	http://gov
66	66	Abel	1414	1424	682	D	Be	10170	Berlin-Mitte	
67	67	Abel	1769	1812	953	ev	D	16775	Sonnenberg	http://gov
68	68	Abel	1838	1838	1317	ev	D	27801	Brettorf	http://gov
69	69	Brun	1600	1800	296	D	MV	18273	Güstrow	http://gov
70	70	Bollmann	1600	1800	296	D	Ni	27318	Hoya	http://gov
71	71	Hopp	1700	1900	3394	D	SH	23779	Neukirchen	http://gov
72	72	Leitmeyer	1800	1900	3394	RUS	OPR		Königsberg	
73	73	Geduhn	1800	2000	3394	RUS	OPR		Tawe/Labiau	
74	74	Schröder	1800	1900	3394	D	SH	22880	Wedel	http://gov
75	75	Wolgast	1700	2000	3394	D	SH	22941	Elmenhorst	

Anreicherung mit Geolocation

- Location nicht normiert, insb. Ausland
- Koordinaten (Geolocation) irgendwie “ran-joinen”

ID	FAMILYNAME	BEGIN_DATE	END_DATE	SUBMITTER_ID	DENOMINATION	COUNTRY	REGION	POSTALCODE	PLACE_NAME	PLACE_UF
53	53	Schröder	1800	1900	1199	D	SH	23701	Roge	
54	54	Holling	1700	2000	1199	D	SH	24594	Tappendorf	http://gov
55	55	Barg	1800	1900	1199	D	SH	24594	Tappendorf	http://gov
56	56	Stabe	1800	2000	1199	D	SH	24802	Bokel	http://gov
57	57	Grage	1700	1900	1199	D	SH	24802	Bokel	http://gov
58	58	Aal	1666	9999	1318	D	He	64404	Bickenbach	http://gov
59	59	Aalst, van	1898	1930	798	NL				
60	60	Aart	1727	1799	528	D	Ni	49832	Freren, Krs. Emsland	
61	61	Abbentheren	1716	1743	331	lu	D	27793	Wildeshausen	http://gov
62	62	Abbenthern	1828	1912	1435		D	30159	Hannover	http://gov
63	63	Abeck	1860	1887	409	rk	SL	66740	Neuforweiler	http://gov
64	64	Abehus	1783	9999	331		Ni	27793	Wildeshausen	http://gov
65	65	Abel	1702	9999	682		Ni	30159	Hannover	http://gov
66	66	Abel	1414	1424	682	D	Be	10170	Berlin-Mitte	
67	67	Abel	1769	1812	953	ev	BB	16775	Sonnenberg	http://gov
68	68	Abel	1838	1838	1317	ev	D	27801	Brettorf	http://gov
69	69	Brun	1600	1800	296		MV	18273	Güstrow	http://gov
70	70	Bollmann	1600	1800	296	D	Ni	27318	Hoya	http://gov
71	71	Hopp	1700	1900	3394	D	SH	23779	Neukirchen	http://gov
72	72	Leitmeyer	1800	1900	3394	RUS	OPR		Königsberg	
73	73	Geduhn	1800	2000	3394	RUS	OPR		Tawe/Labiau	
74	74	Schröder	1800	1900	3394	D	SH	22880	Wedel	http://gov
75	75	Wolgast	1700	2000	3394	D	SH	22941	Elmenhorst	

Am Beispiel Goethe



FAMILYNAME	BEGIN_DATE	END_DATE	SUBMITTER_ID	DENOMINATION	COUNTRY	REGION	POSTALCODE	PLACE_NAME
Goethe	0	9999	16529		D			
Goethe	0	9999	16533		D			
Goethe	0	9999	16542		D			
Goethe	0	9999	16553		D			
GOETHE	1604	1633	2911		D	TH	99706	Berka /Wipper
GOETHE	1657	1694	2911		D	TH	06556	Artern
GOETHE	1710	1782	2911	lu	D	He	60311	Frankfurt a.Main
Goethe	1710	1782	20337	ev	D	He	60311	Frankfurt/ Main
Goethe	1710	1782	18368		D	He	60311	Frankfurt
Goethe	1710	1782	19194		D	He	60311	Frankfurt
Goethe	1752	1858	2046	ev	D	St	39596	Hassel
GOETHE	1777	1777	2911	lu	D	BW	79312	Emmendingen
GOETHE	1789	1832	2911	lu	D	TH	99423	Weimar /Thür
GOETHE	1830	1830	2911	lu	D	RP	54574	Rom /Eifel
Goethe	1832	1832	20337	ev	D	TH	99423	Weima
GOETHE	1844	1	2911		D	OBY	83334	Wien Inzell
Goethe	1875	1	731		USA	WI	5	Wisconsin
Goethe	1881	1912	682		D	Be	10000	Berlin
GOETHE	1883	1885	2911		D	SN	04109	Leipzig
Goethe	1898	1968	731		USA	WI	54305	Green Bay, WI

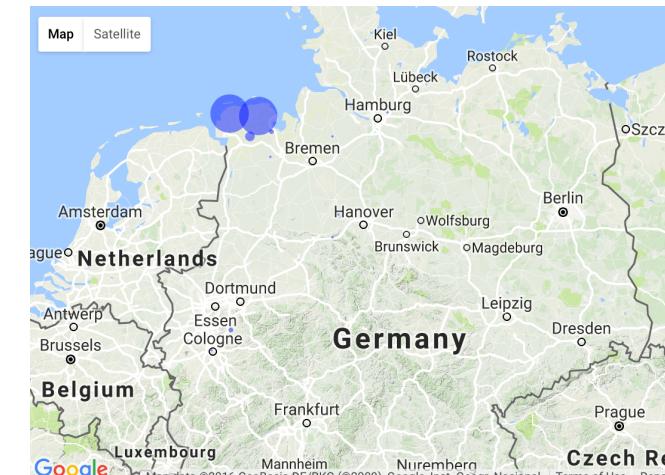
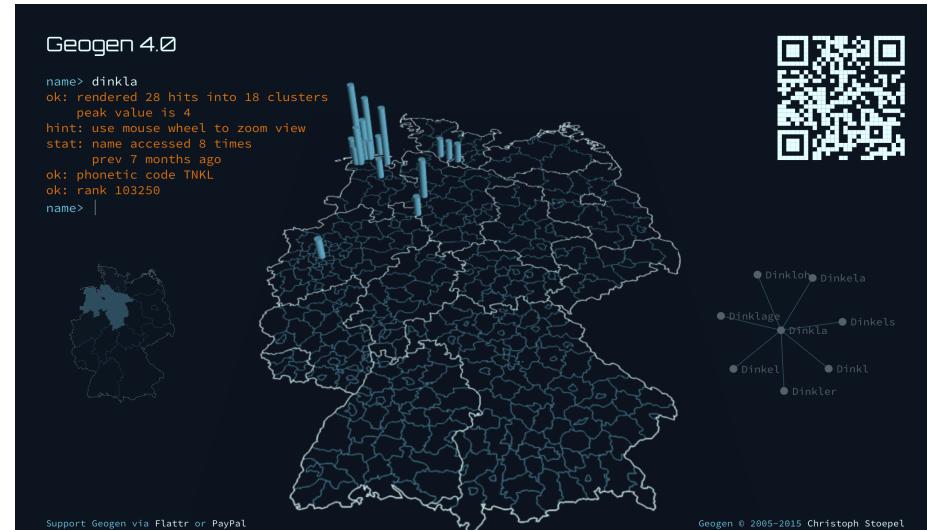
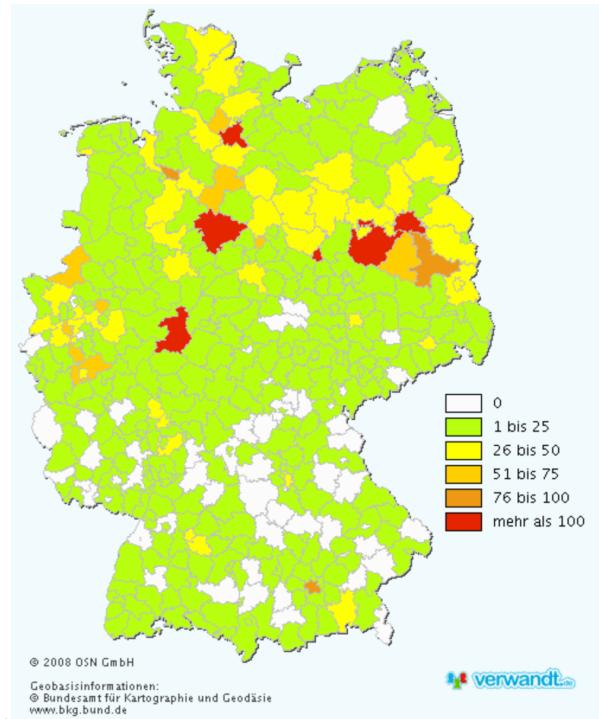
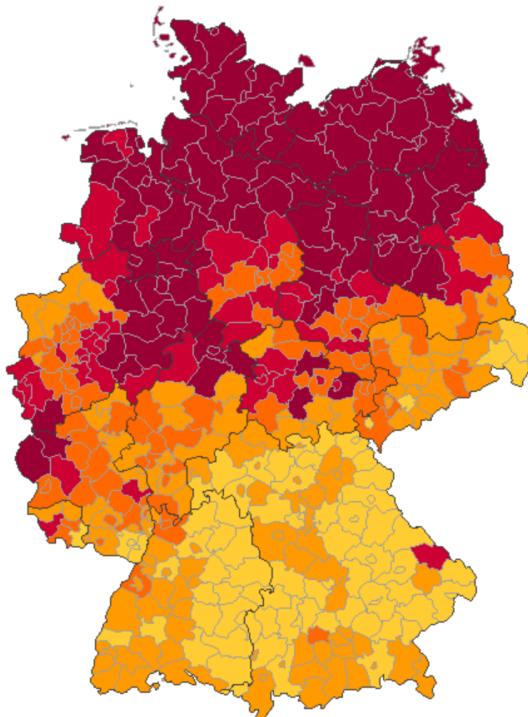
Datenqualität

- Nicht für **wissenschaftliche** Anwendung geeignet (IMHO)
 - Ursache: Händische Eingaben, nicht normiert
 - Doppelte Namen
 - Zeiträume oft grob: 0 – X, X – 9999, X-X
- Bereinigung notwendig
 - Aber “out of scope”

Außerdem gibt es das doch schon ...

- Beispiele
 - <http://www.verwandt.de/karten/>
 - <http://www.ancestry.de>
 - <https://www.myheritage.de>
 - <http://geogen.stoepel.net>
 - <http://www.gen-evolu.de>

Außerdem gibt es das doch schon ...

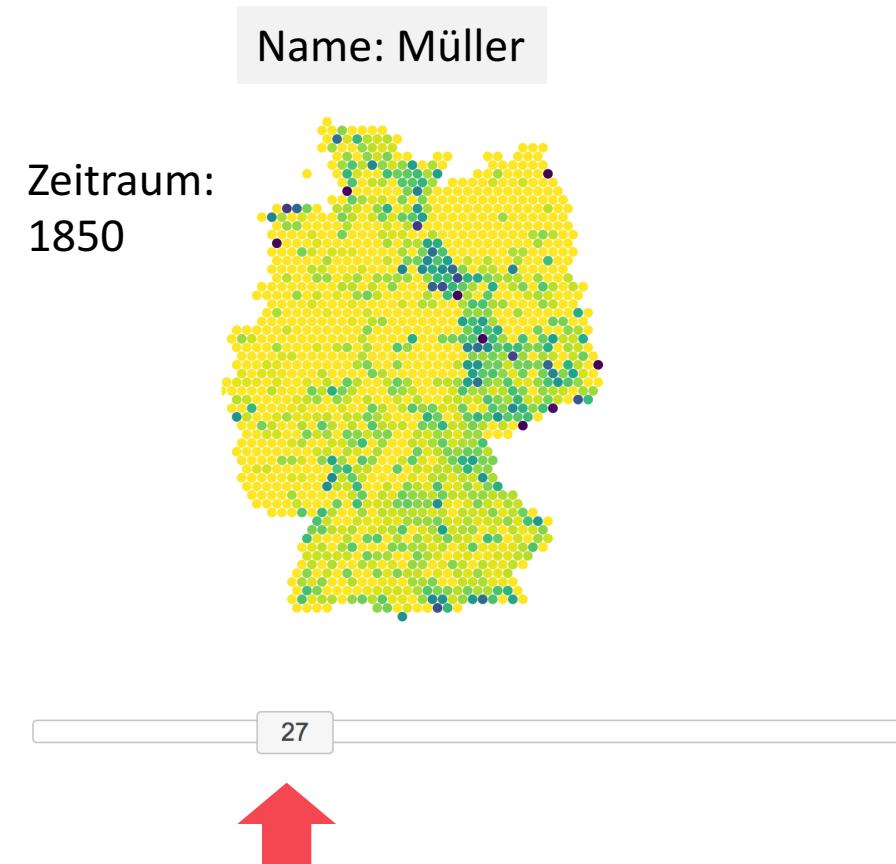


ABER die bisherigen Apps ...

- Aggregieren die Anzahl der Menschen
- Nicht historisch
- Aktuelle Daten oder Snapshots, wie z. B. Telefonbuch
- “Langweilige Karten” und Diagramme
- Keine Open-Source Projekte
 - API und/oder Daten nicht offen

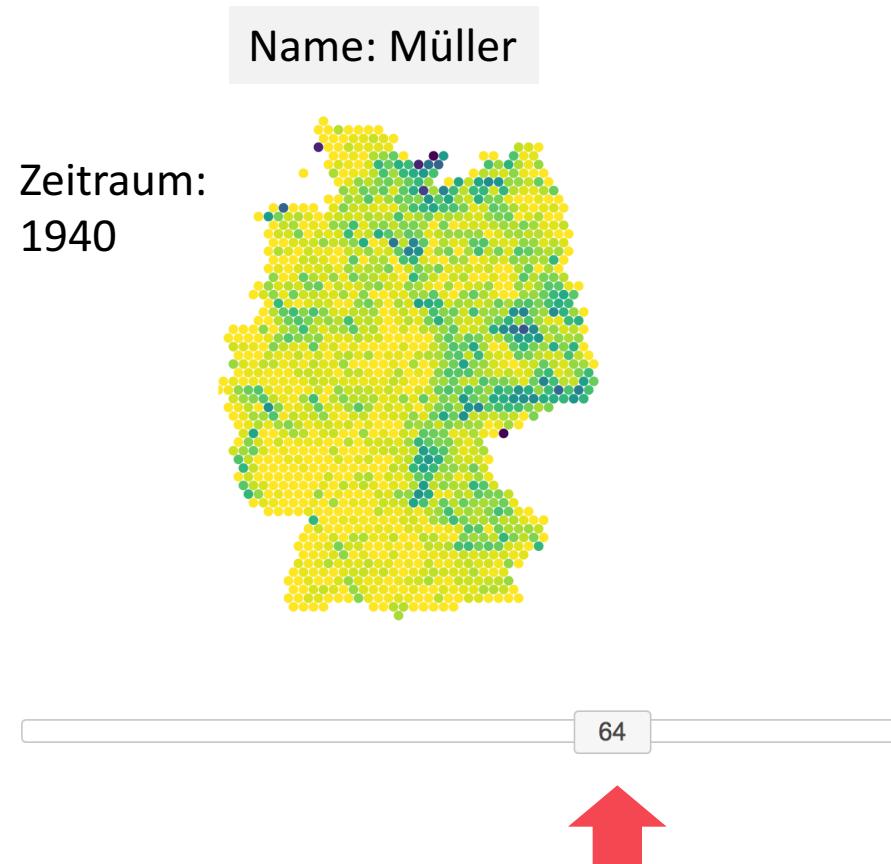
Idee für Visualisierung

- Karte mit Slider für Reise durch die Zeit



Idee für Visualisierung

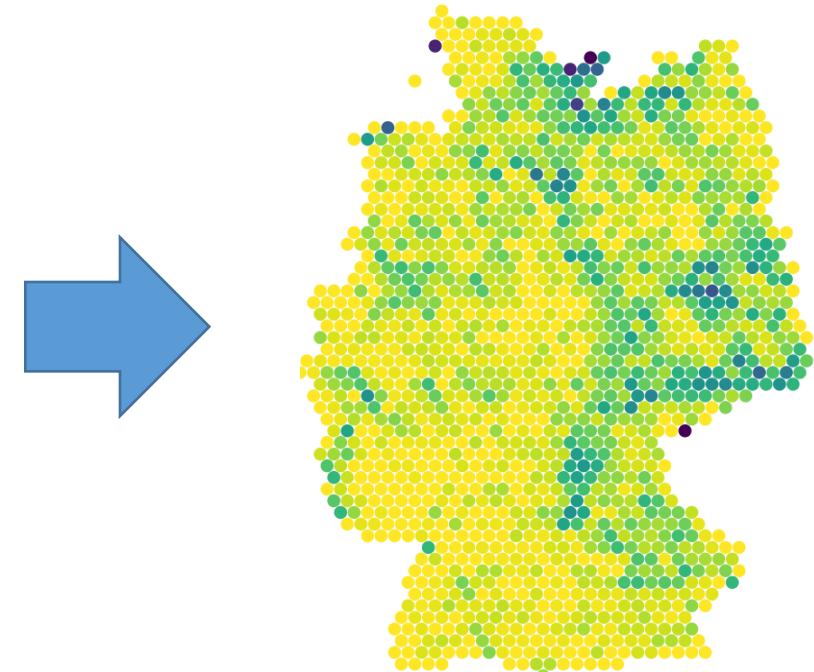
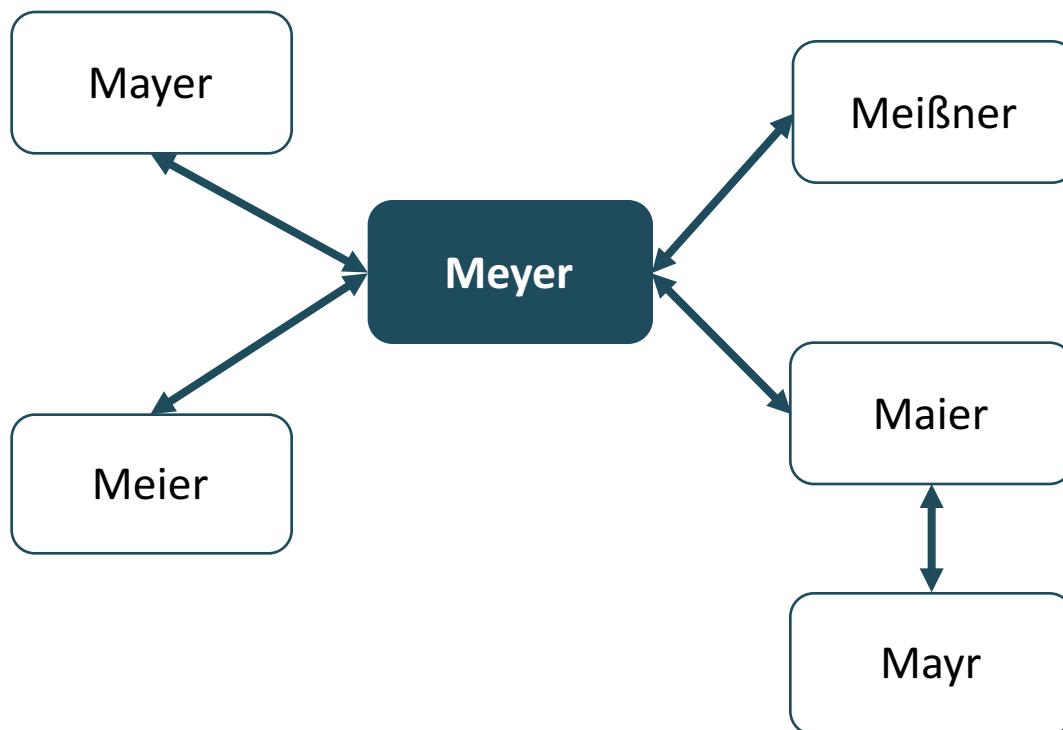
- Karte mit Slider für Reise durch die Zeit



Ähnlichkeiten-Browser

- Ähnliche Namen als Graph, bei Anklick wird Karte angezeigt

Name: Meyer



Zeitliche Entwicklung

■ Zeitleiste

Name: Meyer

Region: D

1650

Meyer

1730

Meißner

1750

Maier

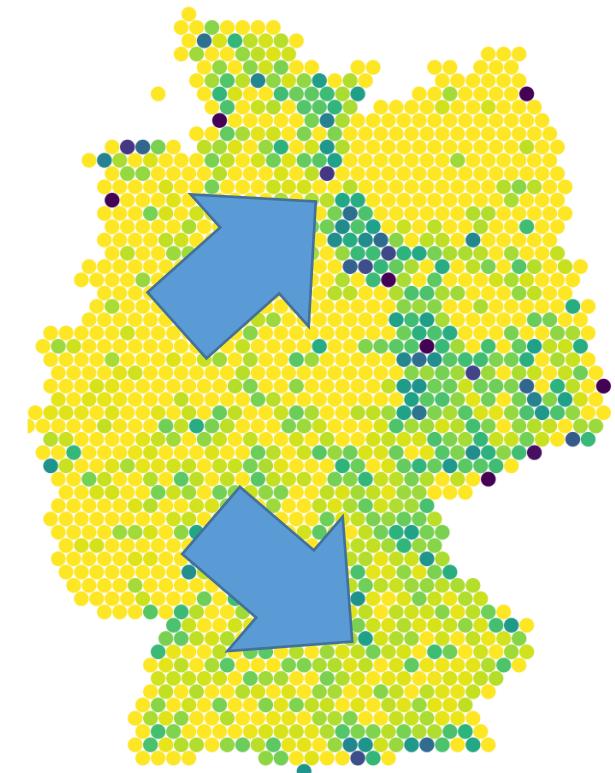
1800

Mayr

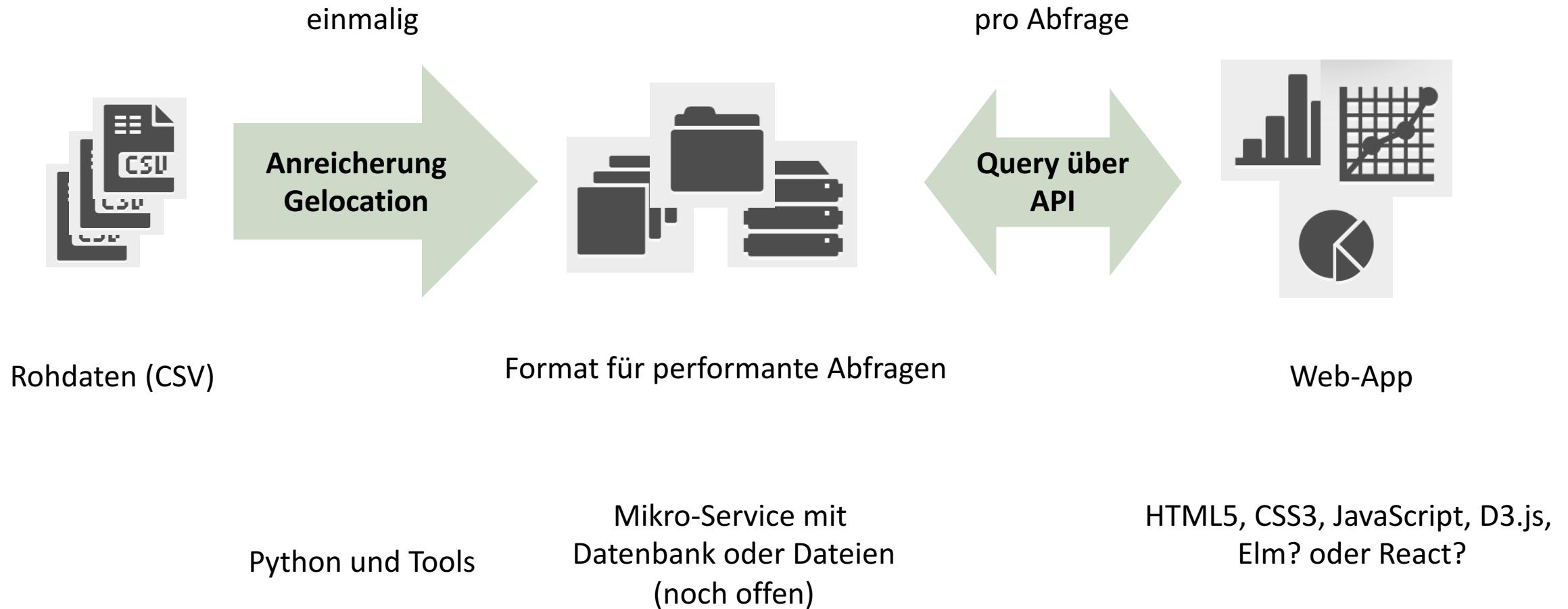


API für Zeitreihenanalyse (Evtl.)

- Analyse von “geometrischen Zeitreihen”
- Beispiele
 - Wanderungsbewegungen
 - Welcher Name war als erster da?
 - Meyer oder Meier?



Architektur



Ziele des Projekts

- Visualisierungen erstellen
 - Slider und Ähnlichkeitenbrowser auf jeden Fall
- Möglichst API extrahieren
 - Damit auch anderere Entwickler Visualisierungen erstellen können
 - Und als Open-Source veröffentlichen
- Optional
 - Guidelines für Umgang mit Daten wg. Datenqualität
 - “data thinking”, “Denken in Daten”