StatBot.Swiss

Dr. Christine Choirat
Data Science Competence Center
Federal Statistical Office

Dr. Christian Ruiz
Statistical Office Canton Zürich
CORSTAT

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

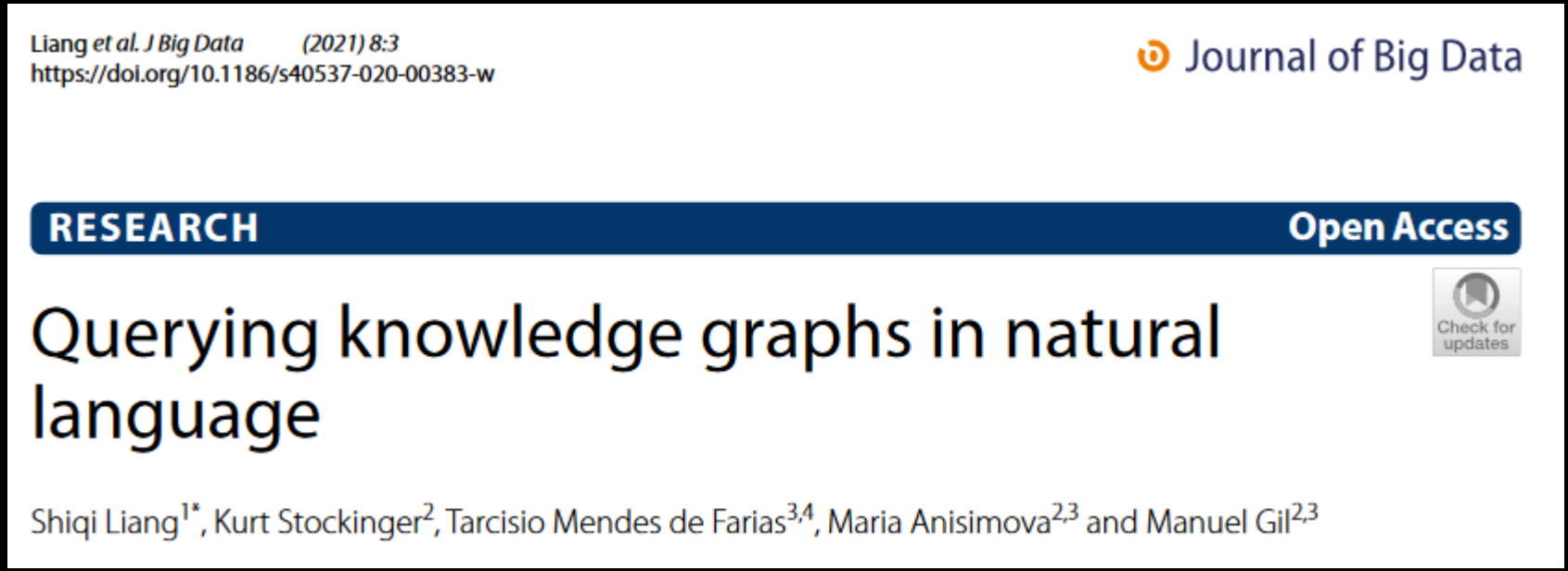
Eidgenössisches Departement des Innern EDI
Département fédéral de l'intérieur DFI
Federal Department of Home Affairs FDHA
Bundesamt für Statistik BFS
Office fédéral de la statistique OFS
Federal Statistical Office FSO

Swiss Statistics Meeting, Lugano, 6 September 2021

Intro

- Simplify for a range of users access to information in OGD
- Difference to "chatty" chatbot
- Open-source, replicable

Intro



Liang *et al. J Big Data* (2021) 8:3
https://doi.org/10.1186/s40537-020-00383-w

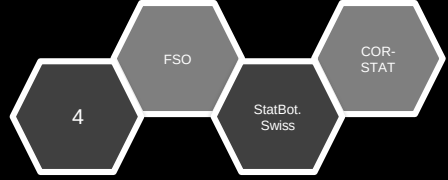
Journal of Big Data

RESEARCH Open Access

Querying knowledge graphs in natural language

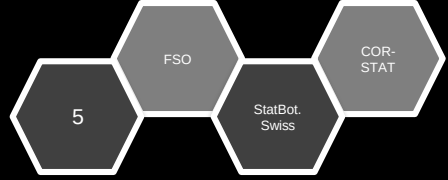
Shiqi Liang[1*], Kurt Stockinger[2], Tarcisio Mendes de Farias[3,4], Maria Anisimova[2,3] and Manuel Gil[2,3]

- Technically feasible



How did we get here?

- Common diffusion project (FSO and CORSTAT)
- NLP-Hackathon at Uni Berne end of March
- 'E-Government Schweiz' project
- Open Government Data and open source



"The way is the aim"

- Final target: MVP of a Statistical bot
  - Feasibility of the endeavour
  - Technical limitations
  - Realistic assessment

- Important objective: Improve common diffusion
  - build competence and knowledge
  - Increase harmonization and standardization

Project Overview

Mid-2022

End-2023

Init. phase → Concept. phase → Implement. phase → Realization phase

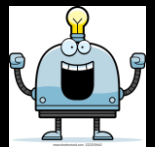
Project Order

Prototype

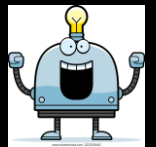
Minimum

MVP

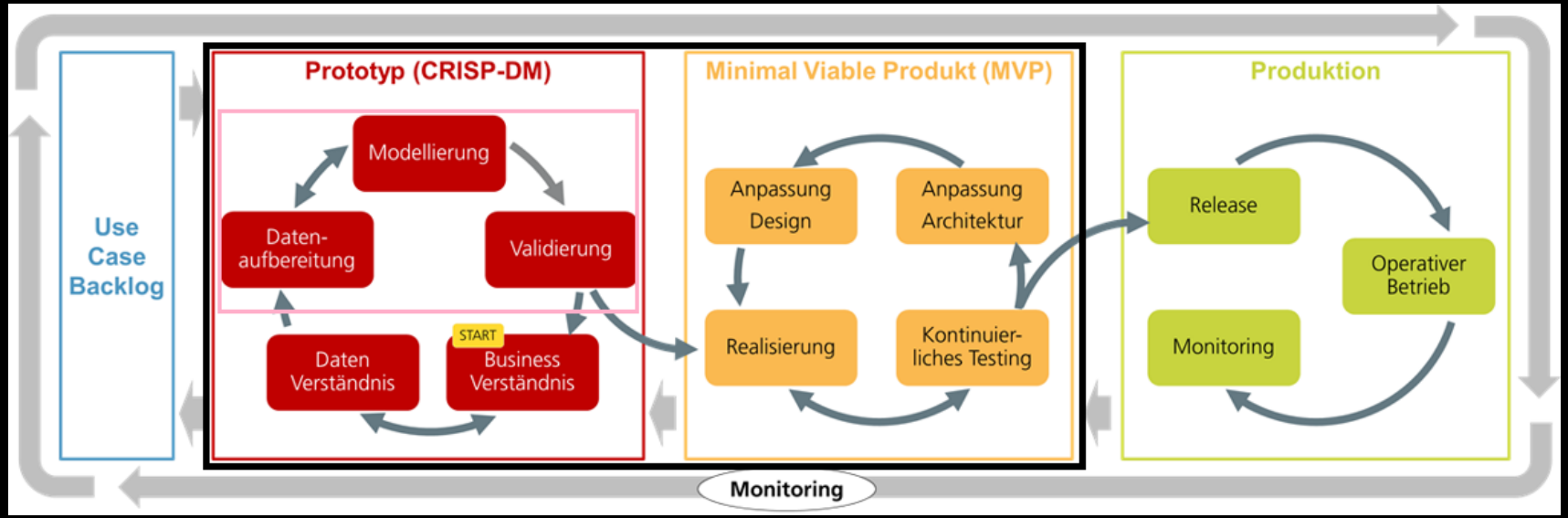
Endurance Test



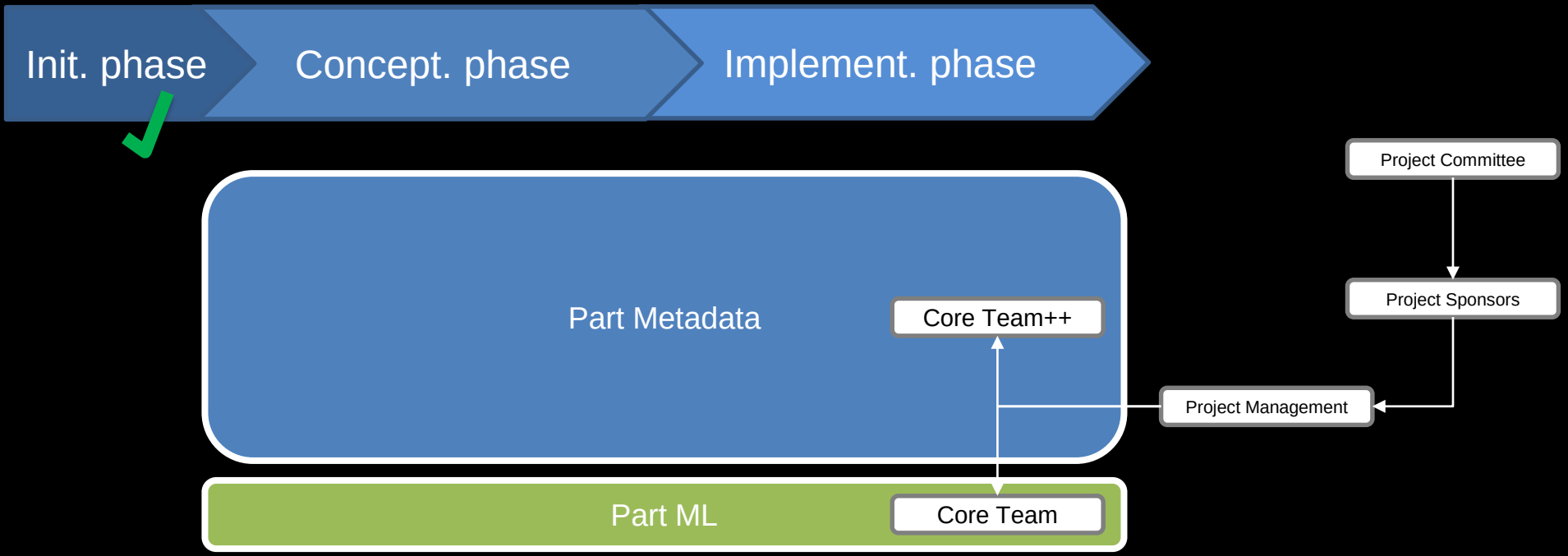
Go Live



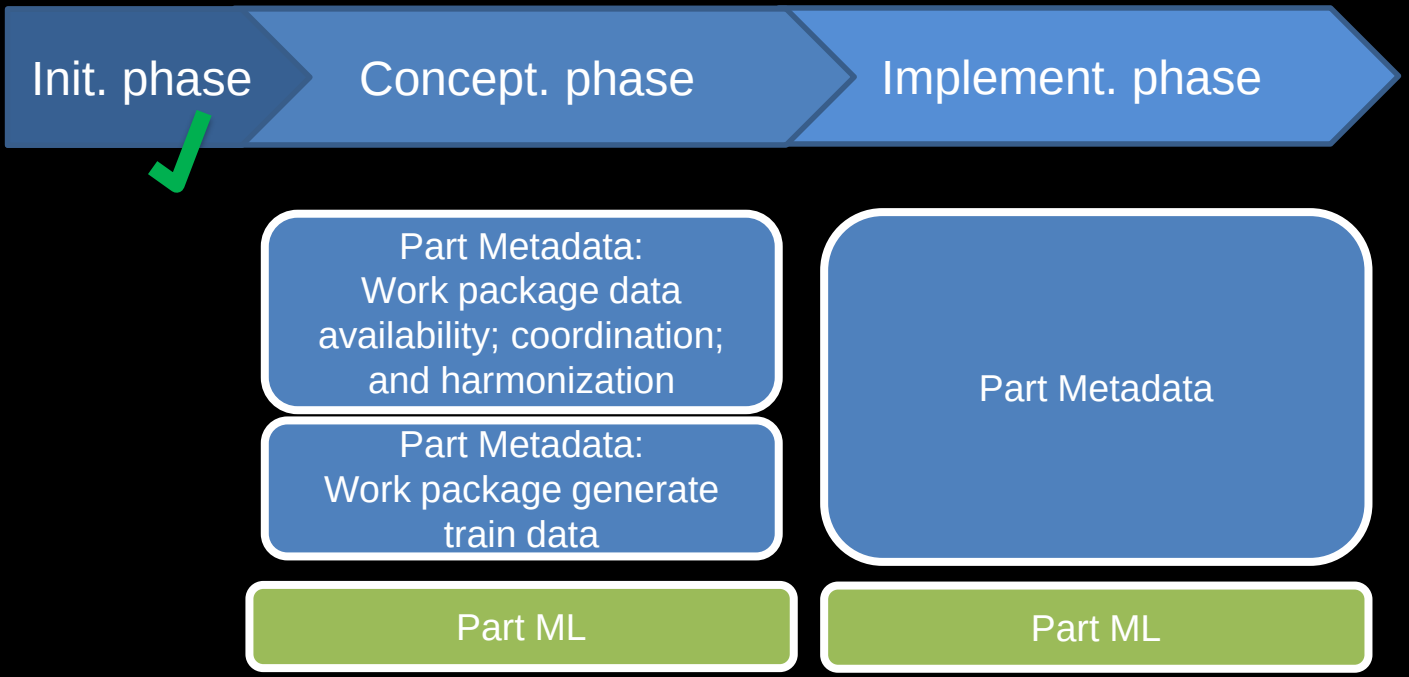
Project overview

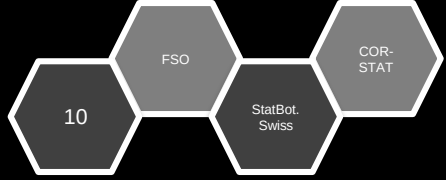


Project Overview



Project Overview





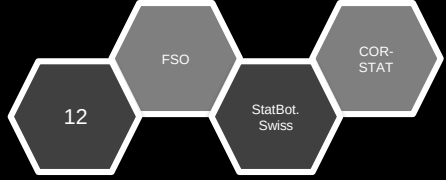
Concrete steps Core Team ++

Part Metadata: Work package data availability; data quality; coordination; and harmonization	Overview/Filtering of datasets of everyone (Mid September)	Map/Matching of datasets (End of October)	Prepare 33% as "raw" LD (End of November)
Part Metadata: Work package generate train data	Expand the fake data generator (End of 21)	Generate gold standard data (mid 22)	Advanced fake data generator (mid 22)

Question-Answer Pairs

	question	sql
1	how high is the share of people living in proximity of a busstop in Elsau?	select iv.value from indicator_values2 iv join
2	how high is the share of people living in proximity of a trainstation in Stäfa?	select iv.value from indicator_values2 iv join
3	how high is the number of cars per capita in Adliswil?	select iv.value from indicator_values2 iv join
4	how high is the share of people living in proximity of a trainstation or busstop in Rüti?	select iv.value from indicator_values2 iv join
5	how high is the share of public transport in traffic movements of people living in Lindau?	select iv.value from indicator_values2 iv join

Script von Thomas Lo Russo: generate_sql_statments_and_questions.Rmd



Conversion to LD

```
filecon<-file("data_ttl/raum_zeit.ttl",open="w",encoding="UTF-8")

writeLines(paste0("<",ts,"property/RAUM> a
<http://www.w3.org/1999/02/22-rdf-syntax-ns#Property> ;\n","<http://schema.org/name> 'Raum'
;\n","<http://schema.org/identifier> 'RAUM' .\n\n"),filecon)

writeLines(paste0("<",ts,"scheme/Gemeinde> a <http://schema.org/DefinedTermSet>
;\n","<http://schema.org/name> 'Gemeinde' ;\n","<http://www.w3.org/2004/02/skos/core#broader>
<",ts,"scheme/Bezirk> .\n\n"),filecon)

writeLines(paste0("<",ts,"scheme/Bezirk> a <http://schema.org/DefinedTermSet>
;\n","<http://schema.org/name> 'Bezirk' ;\n","<http://www.w3.org/2004/02/skos/core#broader>
<",ts,"scheme/Kanton> .\n\n"),filecon)
```

Script: pipeline_DB2TTL.Rmd

See also: LOD01_DCAT2TTL.py

Querying the triple store

	count	place	time	id
1	48.9	"Dietikon"@de	1483138800	Z31122016R001620048XXX0000XXX0000XXX0000
2	17.0	"Birmensdorf"@de	1356908400	Z31122012R001610048XXX0000XXX0000XXX0000
3	11.0	"Henggart"@de	1104447600	Z31122004R000250048XXX0000XXX0000XXX0000
4	43.0	"Lindau"@de	883522800	Z31121997R001220048XXX0000XXX0000XXX0000
5	15.0	"Uetikon a.S."@de	1262214000	Z31122009R001140048XXX0000XXX0000XXX0000
6	5.0	"Kyburg (bis 2015)"@de	1167519600	Z31122006R001210048XXX0000XXX0000XXX0000

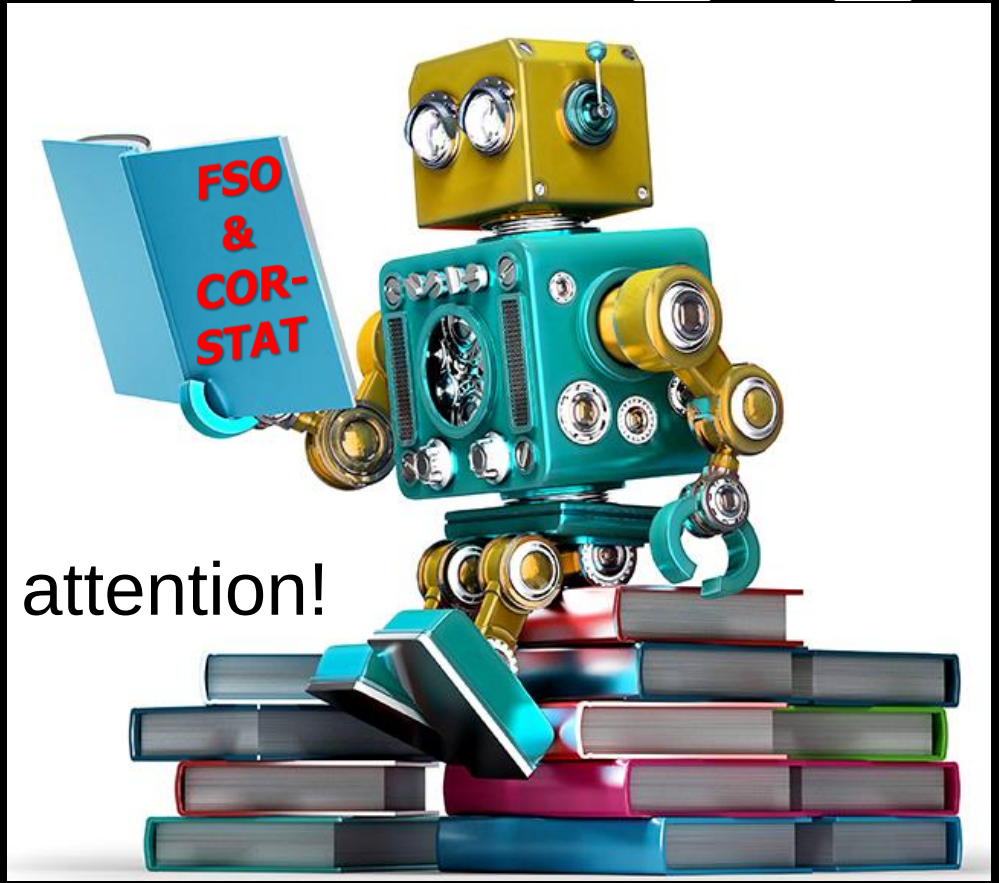
Script: examples_SPARQL/simple_query.R

Next steps

- For you: Are you interested?
  - Join the core team ++
  - Get in contact with us

- For us: Showing that we can achieve the prototype until summer 2022

Thank you for your attention!



Source: CC0 Public Domain with modifications