

Munich Data Science Institute Technische Universität München

Knowledge Distillation From Big Administrative Data Ayshah Chan, Marco Körner

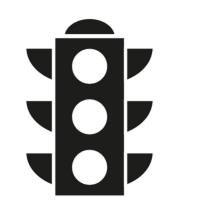
Background

Administrative data offers a huge repository of data that can be used to *further data-driven modelling* but struggles with issues with *data alignment* and harmonization.

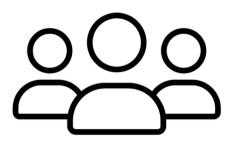
ADMINISTRATIVE DATA







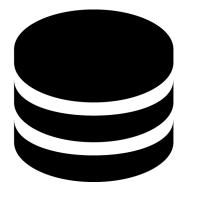
Traffic data



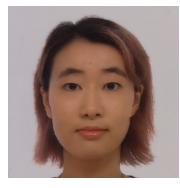
Population data



Migration statistics



Other data collected by public agencies



Ayshah Chan, M.Sc. TUM Chair of Remote Sensing Technology Arcisstraße 21, 80333 Munich www.asg.ed.tum.de/lmf/chan avshah.chan@tum.de

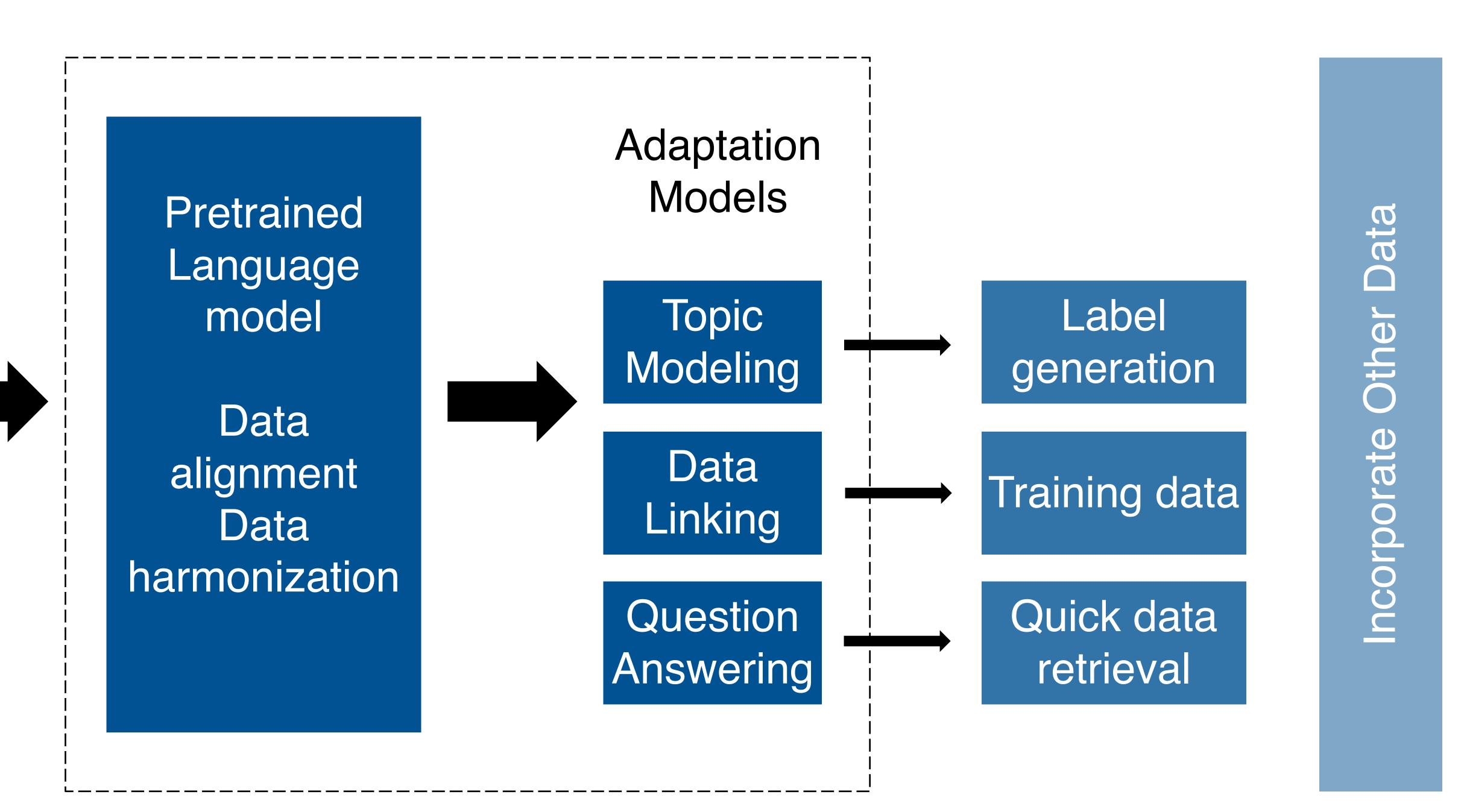


Existing models rely on sequential sentences and whereas administrative paragraphs

Foreseen Challenges

tabularized short phrases and numbers. Significant computation requirements call for *more efficient training* methods.

FOUNDATION MODEL



Prof. Dr. rer. nat. habil. Marco Körner **TUM Chair of Remote Sensing Technology** Arcisstraße 21, 80333 Munich www.asg.ed.tum.de/Imf/koerner marco.koerner@tum.de

MDSI Opening + General Assembly, September 14, 2023

data contain

Expected Outcome

TASKS

A *foundation model* trained on *administrative data* that will enable the *creation of* a wide variety of *training data* and *further applications* in different domains.



APPLICATIONS

Crop Classification

> Traffic Prediction

Data Chatbot