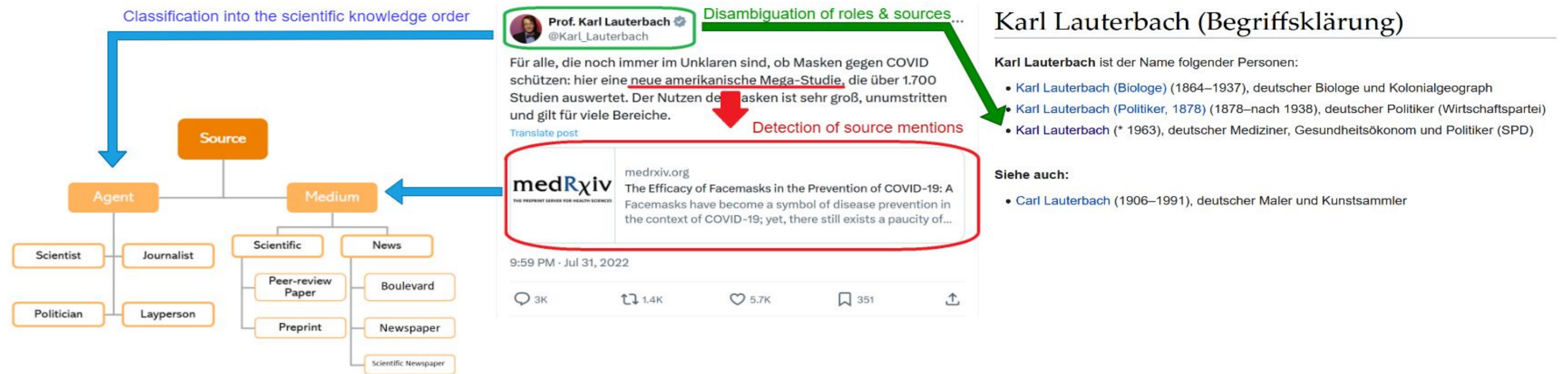


Disambiguation & classification of sources & roles in online science discourse

Philipp Meier Katarina Boland Laura Kallmeyer Stefan Dietze
 Institut für Informatik/Institut für Linguistik, Universitätsstraße 1, 40225 Düsseldorf,
 <firstname.lastname>@hhu.de



“Online Science Discourse”

- Discussion of scientific topics in a fast paced environment like Twitter or Facebook
- Many actors can take part in online discussions, for example scientists, politicians, journalists or laymen


Increasing dietary magnesium intake is associated with a reduced risk of stroke, heart failure, diabetes, and all-cause mortality.

The study findings suggest that increased consumption of magnesium-rich foods may have health benefits.

#Magnesium saves lives <https://t.co/K0M6QdjWcc>
<https://t.co/FUtWZ8jADM>

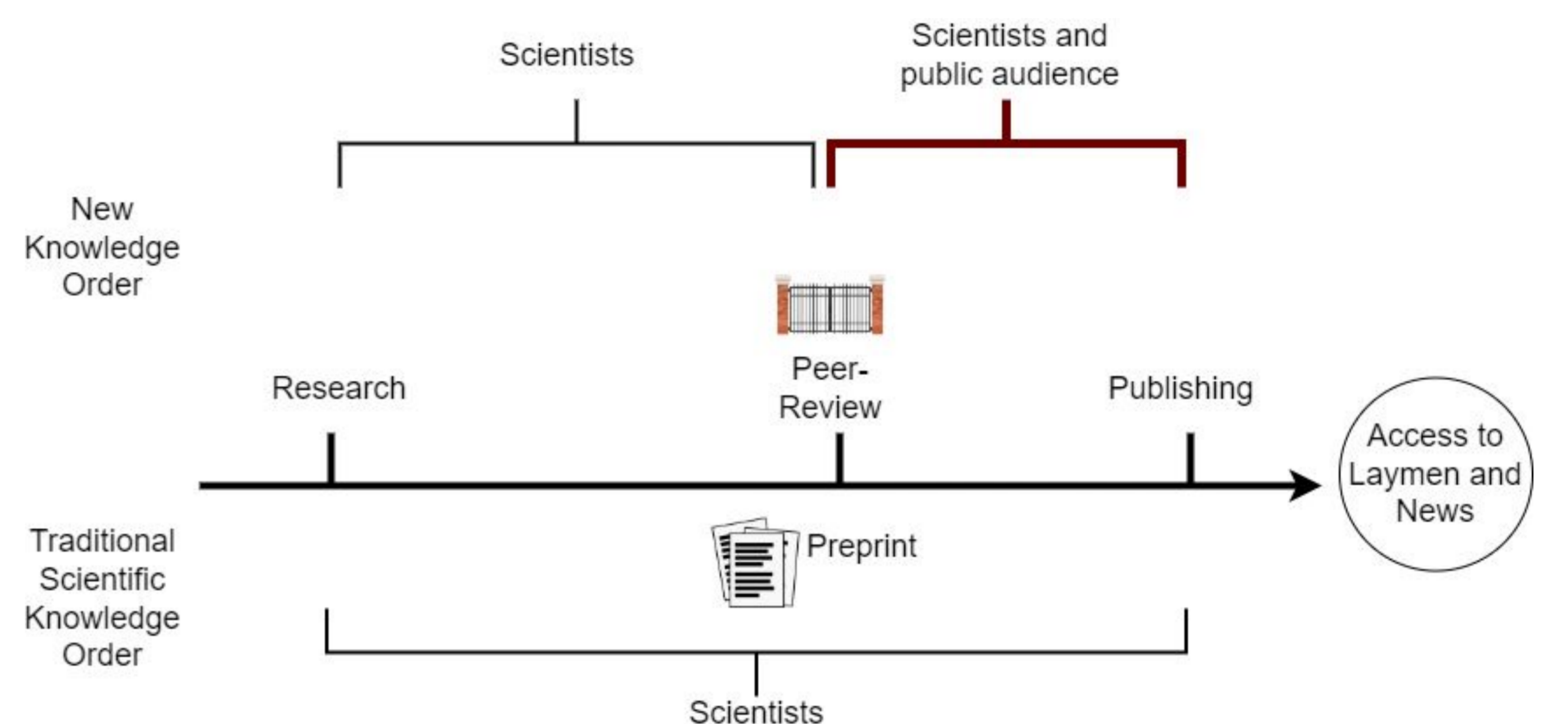
<https://aclanthology.org/2022.emnlp-main.117.pdf>

Data

- TweetsCov19 by GESIS, consisting of ~ 41 Million tweets of 12,8 million users from October 2019 until August 2022
- News Articles, Altmetric Links 

The problems → New Order Project !

- Risks from oversimplification (accidental or deliberate), generalization or instrumentalisation
- **The traditional scientific knowledge order is disrupted**
- **New Order Goal: Understand the disruption and its effects in times of crisis (on the research process, on society)**



Contributions

- WP4 Goal:
 - Classification and disambiguation for sources (paper, news, scientists, politicians) and their roles in the discourse
 - Annotation of language cues (for example ‘Emotionality’)

1. Classification of Language Cues

- Language cues reveal
 - cognitive state and characteristics of the author
 - biases and agendas
- Task may already reveal possible differentiations between users
 - E.g laymen may use more emotional and subjective language than scientists
- Features from Hate Speech and Fake News Detection as well as lexical and grammatical features

3. Disambiguation of roles & sources

- We need to know who is who
 - Check if Twitter profiles have a Wikipedia entry, Research Gate profile or Twitter verification
- Informal references complicate the disambiguation of sources
 - Standard procedure: Link the source mention to its URL
 - Idea: Use the detected claim of the informal reference as input for a search machine
- Attribution of claims allows to cluster users who behave similarly → role

2. Detection of source mentions

- Find science related claims in Twitter data
 - State of the Art (SOTA) uses language models like BERT for claim extraction, e.g SciTweets
- A mention should contain a reference to a source in the ideal case
- Difficult: Informal references
 - ‘A study last month claimed that...’

4. Source hierarchy

- Finally, classify sources and roles in the hierarchy
- Task 1 should already support the agent classification task
 - Semi-supervised task: Define a seed set of users with clearly defined roles
- Source classification relies on Altmetrics, metadata from preprint servers and political orientation of citing news articles