

# BLM - meeting 10/2

## Tankar från forskare:

- Matcha nobelpristagare mot **när**/var de omnämns i BLM
  - Hitta alla författare (namn i texten)
  - **Matcha författarna** mot *Wikidata* (*Jacob kanske har något här!*)
    - Wikidata is ok, but have duplicates (libris same)
  - Vilken roll har författaren i sammanhanget
- Topic model kopplat mot chefredaktörer
  - Metadata: chefredaktörer
- Mäta litteraritet/litteraturvetenskapliga termer
  - PoS-tagging på texten och kör mot år/chefsredaktör/etc.
  - TF-IDF, räkna ord och dess förekomst under de olika åren exempelvis
- **Recensionslängd**
  - Hitta alla recensioner, mät antal ord
- **Vem recenserar vem?**
  - Kön exempelvis
  - Kopplat till akademien och nobelpriset
  - kopplat till ålder på författarna som recenserar/recenseras
- **Hitta alla Essäer** (Essai, Essay) och analysera Berättelsemönster
  - Markera upp alla essäer
  - ordklasser (PoS-tag), antal ord, TF-IDF
  - Topic model
- Hitta alla "Litteraturbrev"
  - samma analys som på essäerna
- Skribenters varande i tid
- Kön/Genus kopplat till sättet hen skriver sina texter
- Omslagen:
  - Snabbfilm av frontpage? (NewYork times)
  - Textanalys på omslagen
  - visuell analys
    - foto
    - rökare
- Mata en modell med text som sedan skriver en essä kopplat till tiden?
- Förekomst av andra språk

## Annan bra data:

- Finns ett register varje år (första numret tror jag) där metadata finns, skulle den gå att använda?

## Metadata:

- Utgivningsår
- Författare
- Formgivare
- Chefredaktörer

# The Nobel Prize as a data-generating process

