

28 November 2019

# Big data pilot for performance audit on EU agencies



EUROPEAN  
COURT  
OF AUDITORS

# Organisation and Data

- Three months pilot: July (data collection), August (testing), September 2019 (reporting)
- **Over 200 000 news articles** gathered in cooperation with JRC
- 43 European agencies
- Audit time frame from January 2018 to July 2019 (1,5 year)
- 5 488 distinct sources from all around the world
- 169 distinct ISO countries
- 66 distinct languages



# Three types of big data analyses

## Part 1: Summary statistics (completed)

*What impact could agencies have in the media? (audit question on “external scrutiny” and “accountability”)*

- Total and average news per agency
- News duplicates
- News outliers + event detection through text mining
- Geographical distribution of news sources per agency

## Part 2: Network analysis (completed)

*What are the relations between Agency-Agency, Agency-Executive Directors, Agency-EU entities, Agency-Member States (“references”) in the media? (audit question on “cooperation” and “coordination”)*

## Part 3: Sentiment analysis (drop-off with lessons learnt)

*Are news articles positive or negative about agencies?*



# Contact details

**Di Hai**

**Private Office of Alex Brenninkmeijer**

di.hai@eca.europa.eu

**Want to know more about our work:**

Visit our website: [eca.europa.eu](https://eca.europa.eu)

Follow us on Twitter [@EUAuditors](https://twitter.com/EUAuditors)

Check our page on LinkedIn and Facebook

Watch our videos on YouTube:

**EUAuditorsECA**

Contact us at: [ECA-info@eca.europa.eu](mailto:ECA-info@eca.europa.eu)



EUROPEAN COURT  
OF AUDITORS