



# EUROPEAN LANGUAGE DATA SPACE



**Productie, beheer en marktontwikkeling van taaldata:  
welke obstakels overwinnen?**

## Panel 2

Veronique Hoste (UGent) & Carole Tiberius (INT; ULeiden) - Moderatoren  
veronique.hoste@ugent.be; carole.tiberius@ivdnt.org

23-01-2025 LDS Country Workshop voor België en Nederland  
<https://language-data-space.ec.europa.eu>



**EUROPEAN  
LANGUAGE  
DATA SPACE**

## Panel



**Moderatoren: Veronique Hoste & Carole Tiberius**

UGent & INT; ULeiden



**Mark Westerhoff**

CTO, ANP



**Saskia Lensink**

AI consultant, TNO



**Karim Dahdah**

Innovation Project Manager, VRT



**Tom Vanallemeersch**

Language AI Adviser, CrossLang

# Karim Dahdah

Project manager van Vlaamse en Europese innovatieprojecten in de mediasector.

- **Narrate:** conversational AI in mediatoepassingen
- **Solid4Media:** meer controle op het delen persoonlijke gegevens door middel van persoonlijke datakluizen
- **Automatische Ondertiteling voor Vlaamse content:** samenwerking tussen DPG Media, Play Media en VRT rond gezamenlijke ontwikkeling van een speech-to-subtitle-model.
- **TEMS:** Media Data Space
- **CityStory:** user generated storytelling
- **SignON:** ontwikkeling van een multimodal taalmodel om vertaling tussen gebarentalen en auditieve talen te faciliteren



VRT is de openbare omroep in Vlaanderen. We willen iedereen in Vlaanderen verbinden door het vertellen van verhalen die voortdurend informeert en inspireert. Met een open geest, onpartijdig, objectief en met respect. Kortom, we willen iedereen mee.

# Saskia Lensink

- **PhD Linguistics Universiteit Leiden:** dissertatie over opslag versus computatie taal
- **Consultant bij TNO:** gespecialiseerd in taal- en spraaktechnologieën. Grote focus op AI-strategieën voor Europa
- **Product Manager GPT-NL:** Ontwikkeling van een soeverein Nederlandse LLM
- **HORIZON TrustLLM:** Democratiseren van betrouwbare en feitelijke LLM-technologie voor Europa
- Co-lead **BDVA Task Force:** bevorderen van verantwoord en duurzaam gebruik van AI- en datatechnologieën
- **Spreker** over generatieve AI



**TNO** innovation  
for life

# Tom Vanallemeersch

- **Achtergrond:** PhD in Computatieve Taalkunde, (KuLeuven) met expertise in machinevertalingen en meertalige taaltechnologie
- **Huidige rol:** Language AI adviser bij CrossLang (Powerling group Innovation Hub)
- **Europese projecten:** ApeQuest (2018-2020), ELRC (2018-2022), Operas (2023), AI4Culture (2023-2025), LLMs4EU (begin 2025)
- **Data in de praktijk:** data creatie en management, data kwaliteit, low-resource scenarios, synthetische data



**CROSSLANG**

TRANSLATION AUTOMATION

# Mark Westerhoff

- CTO van het ANP (Algemeen Nederlands Persbureau)
- AI lead binnen ANP
- AI projectleider G39 Trusted Newshub (Groep van 11 onafhankelijke Europese persbureaus)



# Debatpunten

## 1. Complexiteit van data

- Welke data nodig?
- Culturele identiteit van Europa?
- Hoe waarborgen we inclusie van verschillende types data? Tekst, spraak, gebaren, ...
- Wat is de minimale kwaliteitseis waaraan data moet voldoen?
- Wat met synthetische data?
  - Hoe maak je synthetische data die kwalitatief goed genoeg is?
  - Wat zijn de alternatieven, bijv. scrapen?

# Debatpunten

## 2. Waardebepaling van data?

=> data als grootste asset bij vele bedrijven: balans tussen openheid en commerciële doelen?

## 3. Waarom is het zo moeilijk om aan data te geraken?

- Juridische bottlenecks (contractbeheer) en mogelijke oplossingen? vb. datapods
- Nood aan mentaliteitsshift?





Common European Language Data Space

**Dank u wel!**



A Common European Language Data Space – funded under contract LC-01936389 with the European Union.