

LANGUAGE DATA CREATION, MANAGEMENT AND SHARING: EXISTING PRACTICES AND CHALLENGES IN [COUNTRY]

Panel Session



SESSION BASICS

- **Objective:** investigate the processes for creating, managing and sharing language data at the national level
- **Aim:** update the Country profile (part of the 2019 ELRC White Paper)
- **Format:** moderated panel discussion
- **Duration:** 30-45 minutes
- **Moderator:**
 - representative of an existing data sharing infrastructure [e.g. ELRC (if native speaker, preferably), ELG, CLARIN, the national Open Data portal] OR
 - local researcher in the field of language resources (e.g. the national language corpus)
- **Panelists:**
 - National Open Data portal and/or Public organisation responsible for Open Government actions
 - Existing language data sharing infrastructure(s)
 - Stakeholder(s) identified for the 2019 ELRC country profile, e.g. organisation with considerable contributions in ELRC-SHARE
 - Legal expert

OUTLINE AND CONTENTS: INTRODUCTION BY THE MODERATOR

- Link to previous sessions: LT and AI in general need data -> language data are a special type of data: language is generated all the time, but to train AI systems with language data we need to collect lots of **digital and processable** data on different languages, domains, genres...
- This has proved to be a far from trivial endeavor
- The main obstacles identified by the ELRC White Paper (2019, [link](#)) have been...
- These issues will be discussed with our panelists -> briefly introduce the panelists

OUTLINE AND CONTENTS: QUESTIONS TO PANELISTS

- What is the frame for sharing language data in [country]?
 - Policy level
 - Is language data (sharing) defined/mentioned in the national AI strategy or language policy?
 - Any important developments on policy-level?
 - Does the national open data portal include any (multilingual) language data in processable form?
 - Legal frame
 - What are the most relevant laws/regulations for data sharing / open data?
 - Is language data explicitly defined and is sharing of public language data encouraged?
 - Institutional level
 - Any specifics about panelists' institutional processes for data sharing
 - What do you think is most efficient? Centralized or per organisation creation and management of translations?

OUTLINE AND CONTENTS: QUESTIONS TO PANELISTS (CONT.)

- What are the main challenges for sharing language data? Please let each panelist name from his/her perspective!
 - Examples that may pop-up: legal issues (specify), issues regarding responsibilities and management of data (specify), lack of training (specify) etc.
- What, from your own perspective, would need to be done to improve the situation and facilitate the sharing of language data? Please let each panelist explain from his/her own perspective! For instance:
 - Improving the perception of the value of language data
 - Overcoming structural challenges
 - Improving digital skills and technical capacities
 - Improving translation processes to enable language data sharing
 - Improving procurement practices for outsourcing translations
 - Adjusting the legal framework to enable language data sharing

OUTLINE AND CONTENTS: CLOSING THE SESSION

- Ask participants to go to <https://ec.europa.eu/eusurvey/runner/ELRCCountryWorkshops> and fill in the country survey. If time permits, allow a few minutes for this exercise and/or ask them to fill it in after the workshop.
- Provide the above link through the Zoom chat.

SOME PRACTICAL HINTS

- The moderator can use the suggested master presentation available in Sharepoint (S8.Language data creation_virtual format.pptx)
- Share the list of questions with the panelists in advance – arrange a preparatory session with the moderator (and the panelists), if possible
- Ideally, the panelists will not use slides
- Ask the moderator and the panelists to turn on their cameras during the discussion
- Please go through the 2019 Country profile in the ELRC White Paper (http://www.lrc-coordination.eu/sites/default/files/ELRC_Conference/ELRCWhitePaper.pdf)

THANK YOU FOR YOUR ATTENTION!

Website: www.lr-coordination.eu

Twitter: @LR_Coordination

Email: info@lr-coordination.eu

