

[Click here to be redirected to the virtual room of the Project Expo.](#)

Project abbreviation: MAPA

Project name: The Multilingual Anonymisation Toolkit for Public Administrations

Project coordinator: Manuel Herranz (Pangeanic)

Project consortium:

- Pangeanic
- Tilde
- ELDA
- SEDIA
- LIMSI
- University of Malta
- Vicomtech

Funding: 950000€

Project duration: January 2020 – December 2021.

Main key words: anonymization, Deep Learning, NLP, Neural Networks, Machine Learning.

Background of the research topic: De-identification may provide the means to share language data while also protecting private or sensitive data by spotting then deleting, obfuscating, pseudonymising or encrypting personally identifiable information. De-identification is typically performed for the purpose of protecting an individual's private activities while maintaining the usefulness of the gathered data for research and development purposes.

Goal of the project: building a multilingual anonymization toolkit for public administrations

Project abstract: MAPA project is funded under the Connecting Europe Facility programme, whose goal is the development of an open-source de-identification toolkit for all official European Union languages. The MAPA anonymisation toolkit relies on Named Entity Recognition and Classification (NERC) techniques using the latest neural networks and deep learning techniques. MAPA counts on a data collection activity to provide the necessary training and testing data for the toolkit development. Data is currently being identified and collected for the 24 relevant European languages. As a part of the project, a connection to eTranslation, the online machine translation service provided by the European Commission will be established to foster the provision of multilingual datasets by public administrations that may in turn improve the coverage and quality of machine translation systems. The toolkit will be publicly available and particularly targeted to public administrations in the health and legal domains, as a result of the specific use cases addressed during the development of the project. <https://mapa-project.eu/>

Publications:

[The Multilingual Anonymisation Toolkit for Public Administrations \(MAPA\) Project.](#) Eriks Ajausks, Victoria Arranz, Laurent Bié, Aleix Cerdà-i-Cucó, Khalid Choukri, Montse Cuadros, Hans Degroote, Amando Estela, Thierry Etchegoyhen, Mercedes García-Martínez, Aitor García-Pablos, Manuel Herranz, Alejandro Kohan, Maite Melero, Mike Rosner, Roberts Rozis, Patrick Paroubek, Artūrs Vasiļevskis, Pierre Zweigenbaum. Annual Conference of the European Association for Machine Translation. 3 november 2020.

The "Multilingual Anonymisation Toolkit for Public Administrations" (MAPA) Project. 5th SwissText & 16th KONVENS Joint Conference 2020

