



# Anonymisation Service

Digital Europe Language Tools

*Christopher Scott - IT Project Officer - DGT*

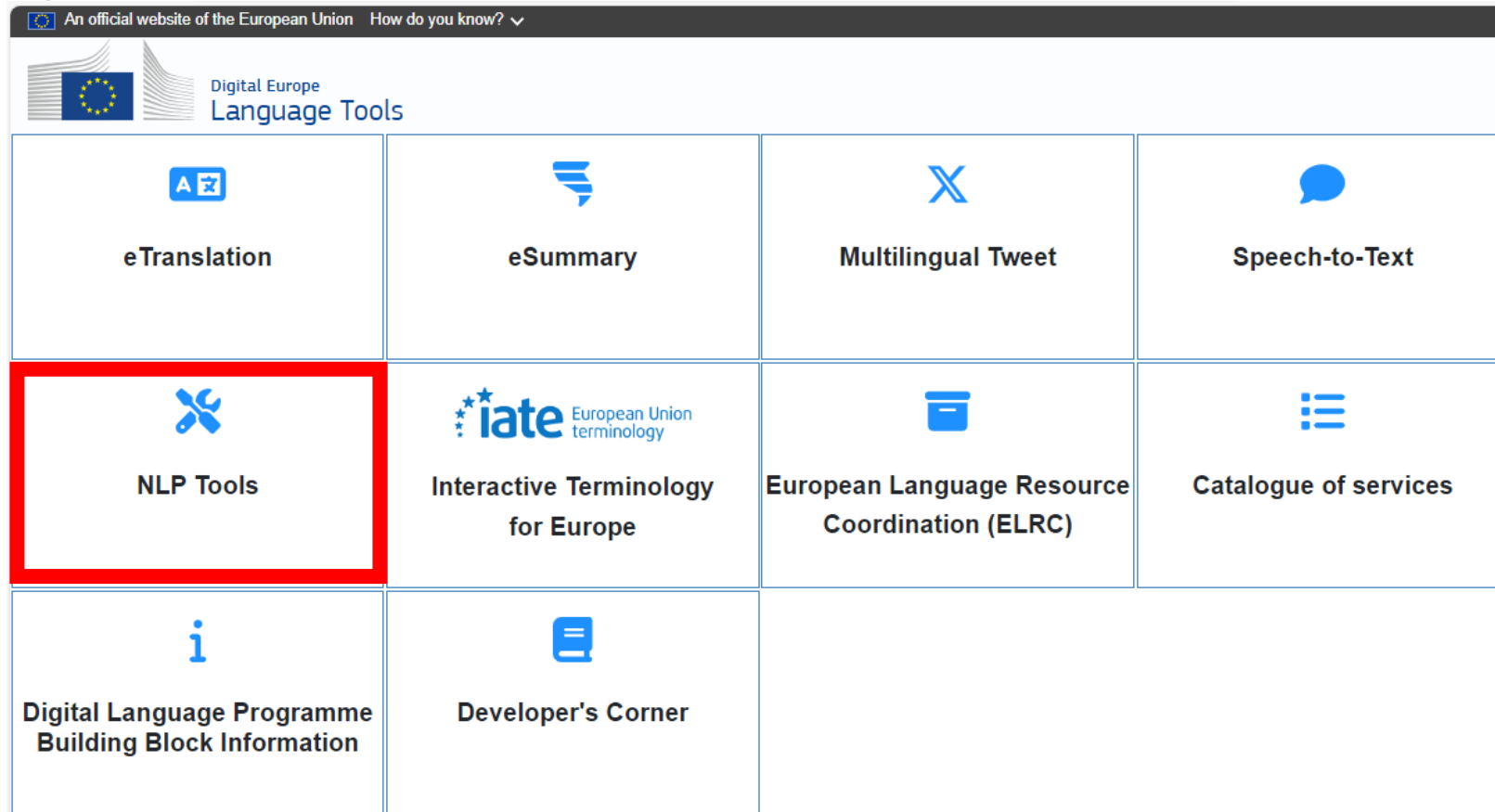
*European Commission*

# MAPA

- **M**ultilingual **A**nonymisation for **P**ublic **A**dministrations
- An AI-based toolkit for anonymisation in all 24 official EU languages
- Developed by private consortium
- Funding through CEF
- Free, open-source

<https://mapa-project.eu/>

# Deployment at the Commission



<https://language-tools.ec.europa.eu/>

# Access to the Service








- European public administrations, local and regional authorities
- SMEs, EU Freelance Translators, Universities
- Non-governmental organizations
- Digital Europe Programme projects

<https://language-tools.ec.europa.eu/>

# Interface – Web (API also available!)

Natural Language Processing Tools My Results My settings English ▼ Logout

Drop files to upload (or click)

Supported formats:       

**From** English ▼  Detect language

**NLP Tools:**

- Anonymization support
  - Obfuscation  Replacement
- Classification
- Named-entity recognition

Submit

# Anonymisation methods

John Smith was born in Dublin on 21/07/1955. He previously worked for Microsoft. He lives in Darmstadt, Germany.

- Obfuscation

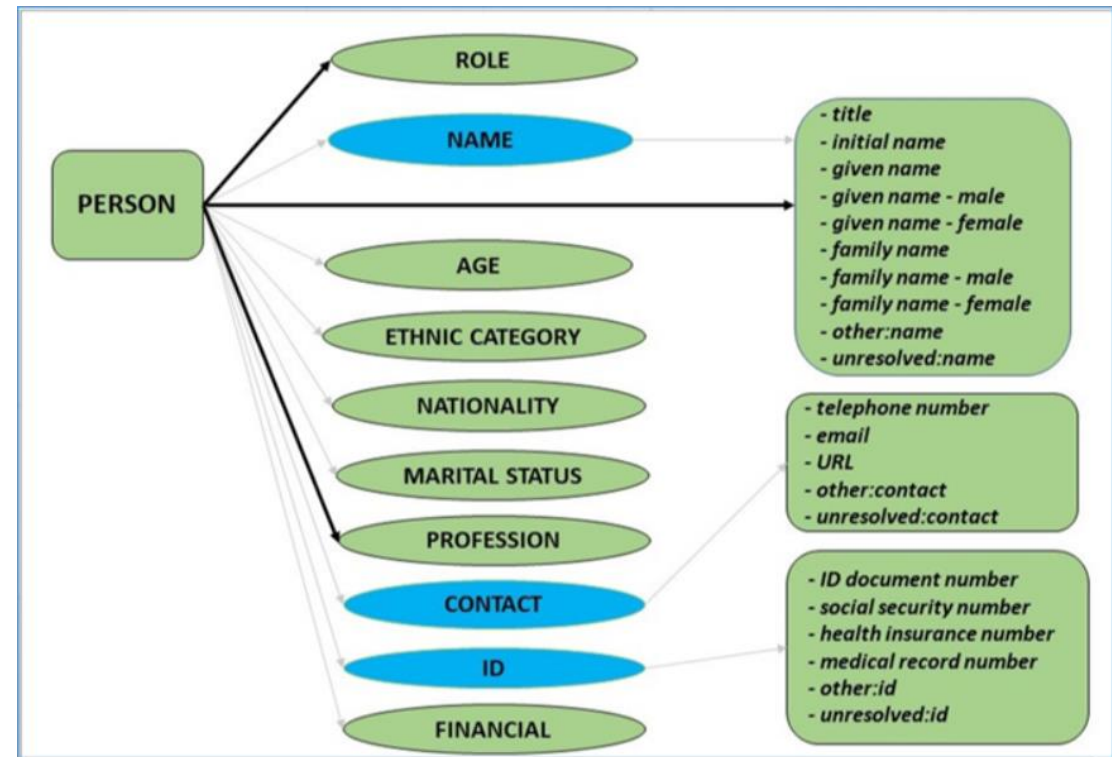
██████████ was born in ██████████ on ██████████. He previously worked for ██████████. He lives in ██████████ ██████████.

- Replacement

**Richard Wilson** was born in **Belfast** on **18/06/1967**. He previously worked for **Google**. He lives in **Nantes, France**.

# Entities for Anonymisation

- Dates
- People
- Addresses
- Organisations
- Amounts
- Vehicles
- Emails
- URLs



# Usage

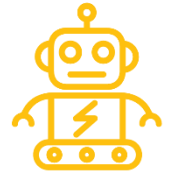
Original text	Blacking out	Replacement
<p style="text-align: center;"><b>MINUTES¶</b></p> <p>Meeting of 31 March 2022, 13.45-16.45¶</p> <p style="text-align: center;"><b>BRUSSELS¶</b></p> <p>The meeting opened at 13.51 on Thursday, 31 March 2022, with <u>Monika Hohlmeier</u> (Chair) presiding.¶</p>	<p style="text-align: center;"><b>MINUTES¶</b></p> <p>Meeting of [REDACTED], 13.45-16.45¶</p> <p style="text-align: center;">[REDACTED]¶</p> <p>The meeting opened at 13.51 on Thursday, [REDACTED], with [REDACTED].¶</p>	<p style="text-align: center;"><b>MINUTES¶</b></p> <p>Meeting of 7 April 2015, 13.45-16.45¶</p> <p style="text-align: center;"><b>LONDON¶</b></p> <p>The meeting opened at 13.51 on Thursday, 7 April 2015, with Michael Koch (President) presiding.¶</p>

Example of anonymisation



# Usage

- Anonymisation of communications from Member States
  - Complaints
  - Requests for information
- Preprocessing carried out by client
- Potential to combine (using APIs) with translation or other services
- 900K requests, 1M documents



# To bear in mind...

- Anonymisation is performed on a best-effort basis
- Results can vary depending on language and entity type
- Confidentiality of data
- Future developments?

# Thank you



© European Union 2024

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

