Challenge sheet

EUSTAT

How could **the standardisation and codification of the postal addresses of external sources be standardised and codified in order to integrate the information in our system**?

Sub-challenges:

- How could an indicator of the reliability of the results obtained by each linked element be incorporated?
- How could the unions with more than one possibility and options with greater probability be displayed for their possible final selection manually?
- How could additional information be used in order to improve the identification and reliability indicators, by consulting with other EUSTAT databases (population statistical register, directory of economic activities, etc.)?

Background:

EUSTAT is the administrative autonomous entity of the Basque Autonomous Community that develops, produces and disseminates official statistical information – that is objective and scientifically proven – for Basque public administrations, social stakeholders and society.

EUSTAT is thus tasked with facilitating the geographical coding tasks for different statistical operations of the institute itself and of other entities, along with standardising those codes. However, data and files are currently received from different external sources with information on postal addresses that have not been codified. A tool is needed to codify the postal addresses based on the standardised codifications generated in EUSTAT in order to be able to integrate the information of those external sources.

Goals:

We are seeking solutions that comply with at least one of these goals, namely to:

- Codify each part of a postal address to the lowest possible level of the hierarchy of the elements that make up a territorial unit for statistics (TUS), taking into account the postal address components (province, municipality, entity/nucleus, street, doorway, floor, number) with the greatest reliability.
- Provide data reliability and in the cases where the reliability is under an agreed scale, be able to view a sub-set of candidates with the greatest possibility, and have the option of selecting one of them.
- Facilitate the use of this solution for the projects of the different EUSTAT areas (such as directories, territory, population).

What are we looking for?

The following aspects will be considered:

- The success and reliability of the codifying must be high.
- The solution must be able to work with other Oracle databases, where the information is stored.
- It must not use excessive resources of the system.
- It must be easily integrated in our system.
- It must be suitable to be used by any type of user.

Process and key dates:

Register using this link to take part in the challenge [still pending].

- ➔ The deadline to receive the response to the proposal is: 23:59 hours on 02/05/2024
- ➔ The semi-finalist startups will be notified on 10/05/2024.
- ➔ The finalist startups will be notified on 240/05/2024.
- ➔ The winning startup will be notified on 13/06/2024.

The selection process consists of 3 phases:

1. **Phase 1:** semi-finalists' pitches with society. Three startups will be shortlisted as finalists. They will take place from 16 to 24 May 2024.
2. **Phase 2:** round of finalists' interviews with the company. They will take place in the week starting 27 May 2024.
3. **Phase 3:** final meeting with finalists to select the winning startup for the pilot scheme with the company. It will be held between 11 and 12 June 2024, during the acceleration programme.

What are the rewards?

Finalist startups (3)

- **2-day govtech** acceleration programme. It will be held in-person in Bilbao on 11 and 12 June.

Winning startup

- **Presentation of the solution at the 2024 BIND GovTech Demo Day**, in-person in Bilbao.
- **Remunerated pilot** up to a maximum of €15,000. The pilot to implement the winning solution will begin in July 2024 and will last 5 months.