

CEFAT4Cities, a Natural Language Layer for the ISA² Core Public Service Vocabulary

Joachim Van den Bogaert, Arne Defauw, Sara Szoc, Frederic Everaert, Koen Van Winckel, Alina Kramchaninova, Anna Bardadym, Tom Vanallemeersch

CrossLang
Kerkstraat 106
9050 Gentbrugge
Belgium

{first.lastname}@croslang.com

Abstract

The CEFAT4Cities project (2020–2022) will create a “smart cities natural language context” (a software layer that facilitates the conversion of natural-language administrative procedures, into machine-readable data sets) on top of the existing ISA² interoperability layer for public services. Integration with the FIWARE/ORION “Smart City” Context Broker, will make existing, paper-based, public services discoverable through “Smart City” frameworks, thus allowing for the development of more sophisticated and more user-friendly public services applications. An automated translation component will be included, to provide a solution that can be used by all EU Member States. As a result, the project will allow EU citizens and businesses to interact with public services on the city, national, regional and EU level, in their own language.

1 Introduction

eGovernment tools and data sets provided by public sector organisations, are typically too fragmented across institutional levels and entities. This is known as the silo effect. To alleviate this effect, the ISA and ISA² (Interoperability Solutions for European Public Administrations) programmes¹ have been supporting the development of interoperability solutions for cross-border and cross-sector public services, for more than a decade. While ISA² Actions, such as the Core Vocabularies and the Core Public

Service Vocabulary Application Profile (CPSV-AP) provide public administrations with templates and tools for describing their public services in a uniform way, one important aspect is still missing, i.e. user-centredness: existing infrastructures are primarily designed to facilitate computer-to-computer interaction, using a common, machine-readable vocabulary. But EU citizens do not speak in standardized data fields. Instead, they use 24 official EU languages to talk to their government: for instance, for an Austrian national living in Belgium, and starting up a new business, it is more important to find an eGovernment procedure in his or her mother tongue, than to know that all data is represented in ISA²-compliant fields.

2 Proposed solution

To speed up the development and adoption of multilingual, cross-border eGovernment services, the CEFAT4Cities project, which runs from April 2020 to March 2022, will build:

- a quality-assured workflow that assists in transforming natural-language eGovernment administrative procedures into multilingual interoperable solutions. CEF eTranslation will be integrated as a tool for developing multilingual resources.
- an open linked data repository containing the results of applying this workflow, i.e. multilingual vocabularies.
- a FIWARE/ORION² extension that allows the context broker to be integrated with the envisaged workflow and resources.

¹ https://ec.europa.eu/isa2/isa2_en, ISA and ISA² support the development of digital solutions that enable public administrations, businesses and

citizens in Europe to benefit from interoperable cross-border and cross-sector public services.

² <https://fiware-orion.readthedocs.io/en/master/>

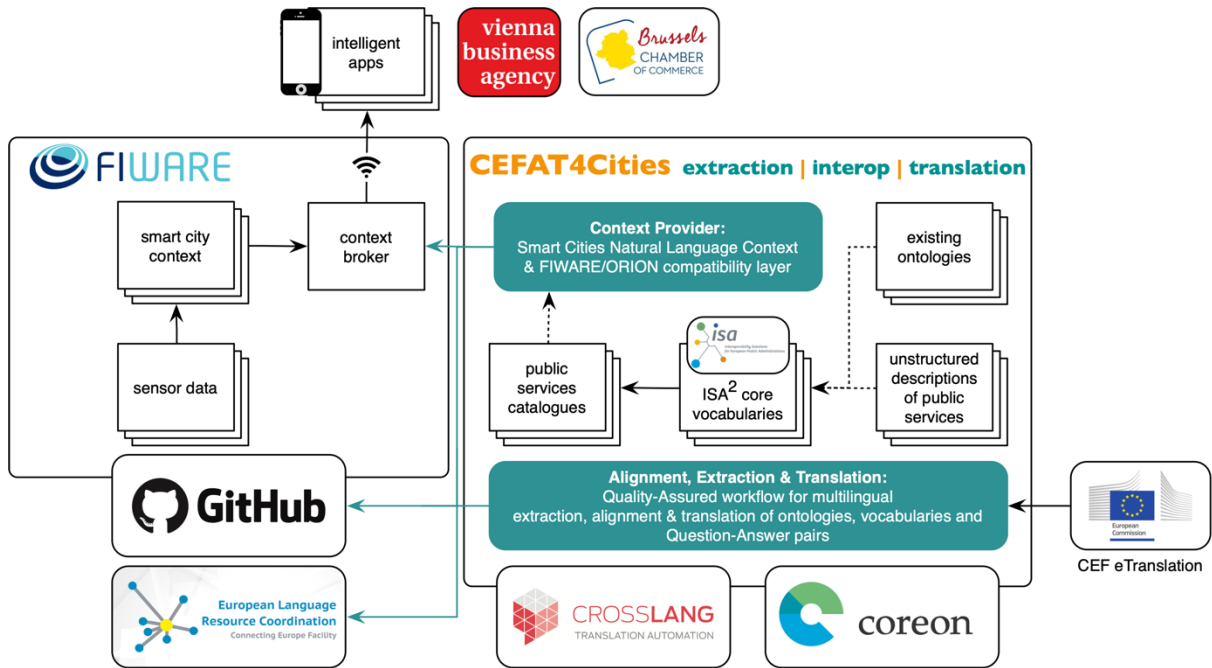


Figure 1: CEFAT4Cities system architecture

To promote the use of the CEFAT4Cities infrastructure within the smart cities community, a dissemination and communication plan, along with a roadmap for future funding, will be developed.

3 Use cases

The effectiveness of the solution will be demonstrated by three use cases:

- CEFAT4Cities will be integrated into the Context Broker Digital Service Infrastructure and the CPSV-AP and OSLO² templates, as a general Natural Language Processing (NLP) layer;
- the Vienna Business Agency will make its public services discoverable in multiple languages;
- the Brussels Chamber of Commerce will create a multilingual wizard for starting a new business in Brussels and showcase how CEFAT4Cities contributes to bringing local public services to a European scale.

The following tangible results will be produced:

- parts of CEFAT4Cities will be distributed as open source software components, enabling the Context Broker to deal with multilingual public service catalogues, through a connection with the CEF eTranslation service.
- the Vienna Business Agency will be able to connect its Content Management System to a multilingual resource, semantically linked across languages.

- the Brussels Chamber of Commerce will have a multilingual app for non-Belgian company founders.
- resources for the following languages will be developed and committed to the ELRC repository: Dutch, English, French, German, Italian, Slovenian, Croatian and Norwegian.

4 Consortium

CEFAT4Cities will be implemented by a consortium of five partners, representing three different countries (Austria, Belgium and Germany), demonstrating a diversified expertise, with access to various stakeholder communities. CrossLang (Belgium) and Coreon (Germany) have a strong track record in NLP, Machine Translation, and Knowledge Management. The FIWARE Foundation (Germany) has plenty of experience in productizing research code and working with the open source community. Vienna Business Agency (Austria) and Brussels Chamber of Commerce (Belgium) have both immediate access to other regions and cities. The targeted users, represented by the cities of Maribor (Slovenia) and Trento (Italy), will involve their networks of problem owners to test the project results and to promote the outcomes.

Acknowledgement

CEFAT4Cities is funded by the EC's CEF Telecom programme (project 2019-EU-IA-0015).