



coreon

Knowledge meets language.

# Building Multilingual Chatbots with Language-Agnostic Knowledge Graphs

Alena Vasilevich,  
NLP Engineer

[alena@coreon.com](mailto:alena@coreon.com)

[www.coreon.com](http://www.coreon.com)



The  
Knowledge  
Graph  
Conference

# Building Multilingual Chatbots with Language-Agnostic KGs

## WHY

CEFAT4Cities: project's background and motivation

## HOW

Chatbot's architecture: Rasa Open Source x Coreon MKS

## LEVERAGING KG

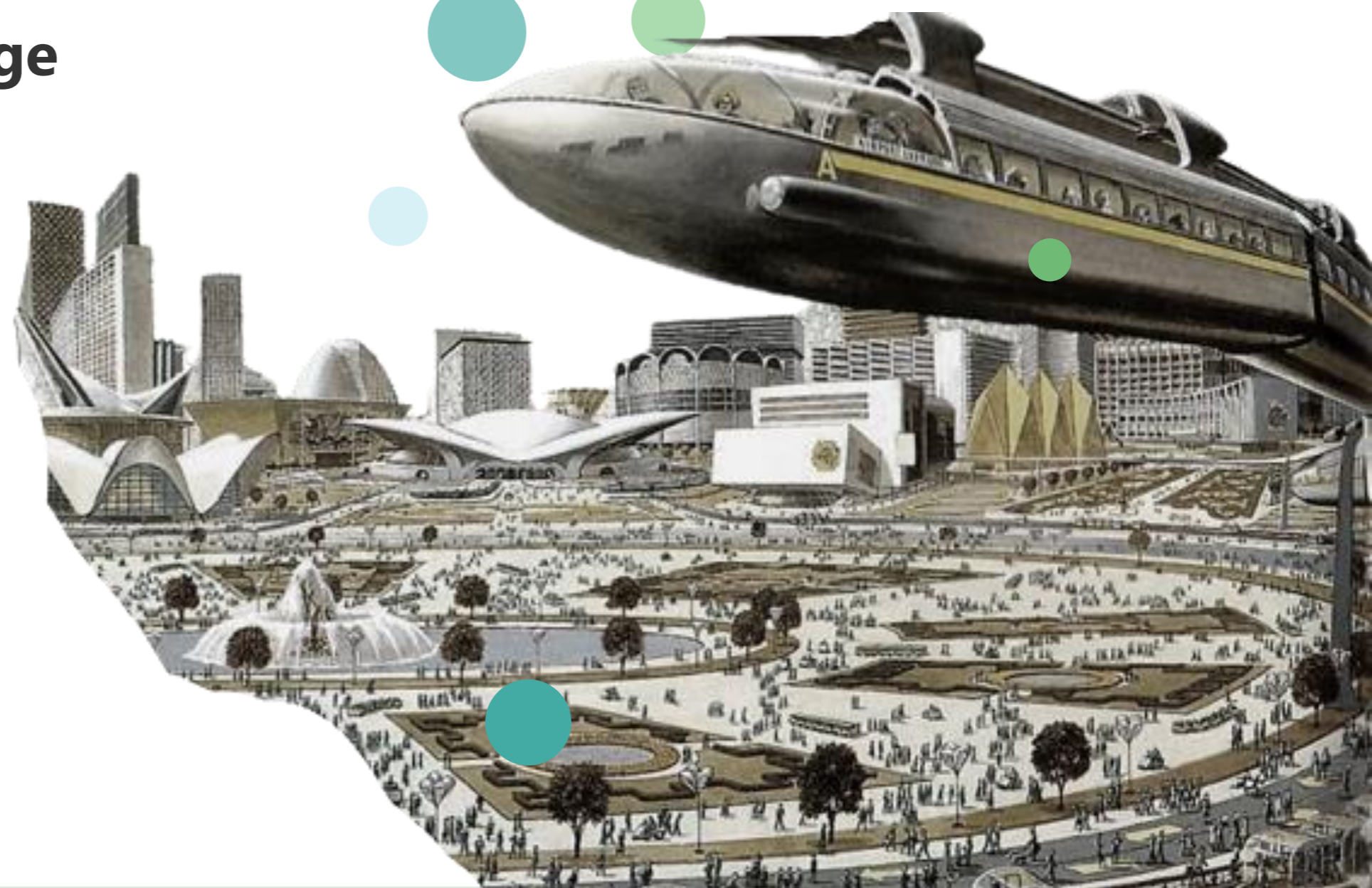
Easy domain knowledge drafting

## ADVANTAGES

Perks: dealing with homonymy & unseen terms; semantic search; agnostic entity management; multilinguality

## **CEFAT4CITIES: Smart Cities Natural Language Context**

- what's it about?
- why are we doing it?



# CEFAT4Cities Action (2020-2022)

- **Targets EU interaction with public services, on a variety of levels**
- **Aims to  development and adoption of multilingual cross-border eGovernment services**
- **Facilitates conversion of natural-language administrative procedures into machine-readable data**
- **Along with achieved results, integrates existing ISA2 and CPSV-AP resources into working real-life applications**

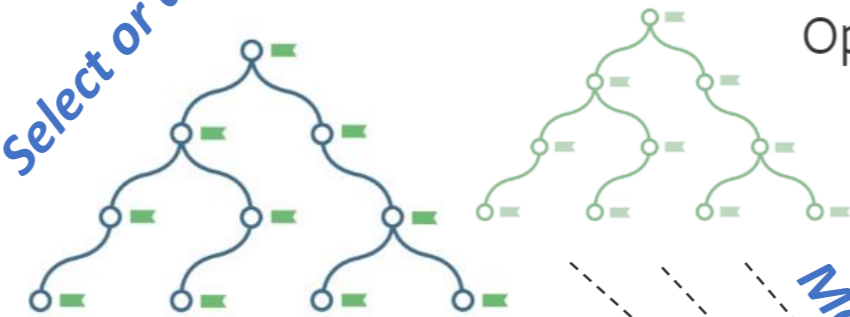


# Data Pipeline: Collect, Curate, & Integrate



Select or discard

Open Standaarden voor Linkende Overheden (OSLO)



Map

Enrich Multi-lingually

Conceptualize, locate, curate



Scan and identify



Share, integrate



Natural language data

Mono- and multilingual candidates

Multilingual Knowledge System(s)

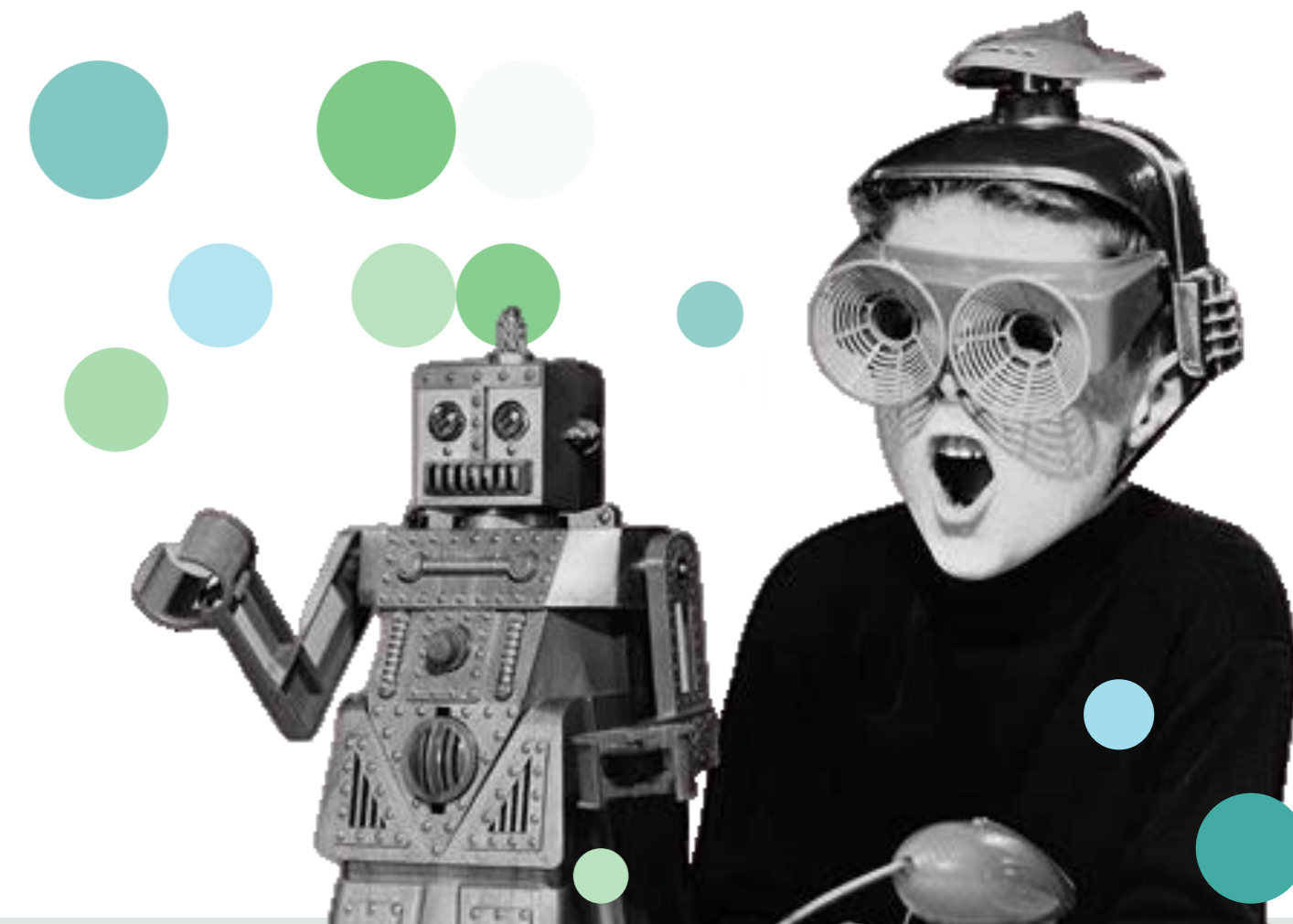
“Consuming” system(s)

Hidden

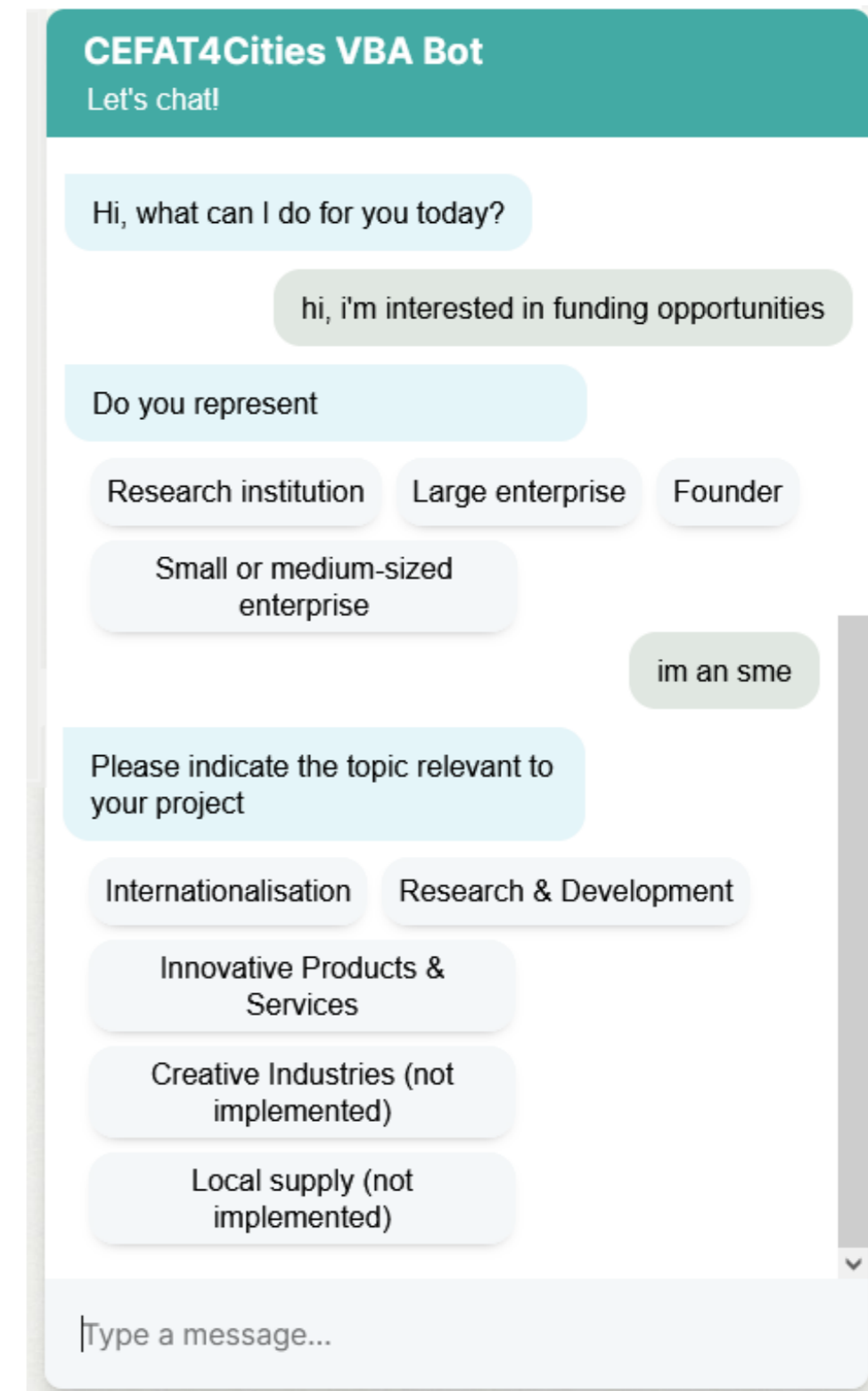
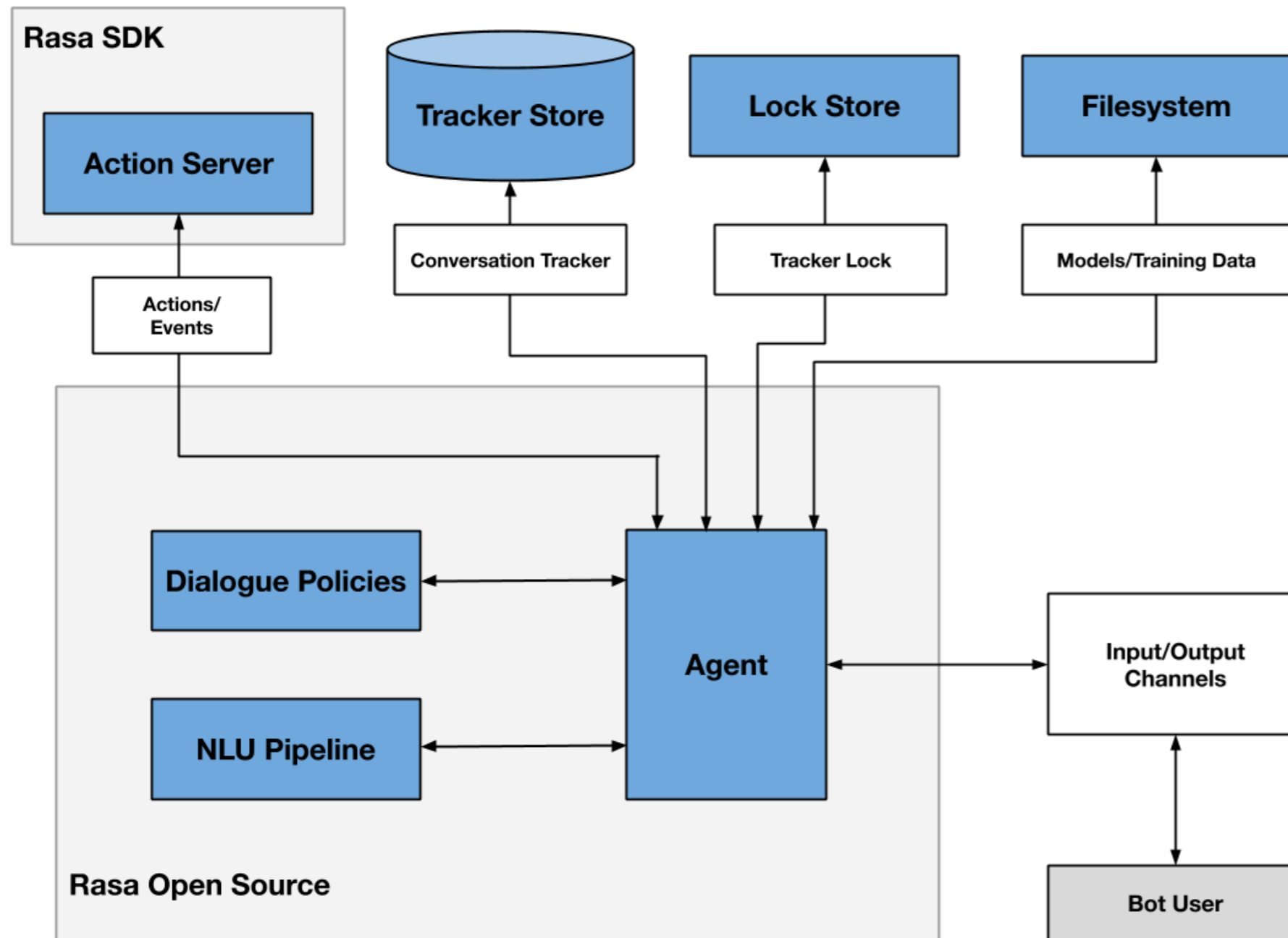
Unsystematized

Systematized, aligned, enriched

## Chatbot's architecture: Rasa Open Source x Coreon MKS

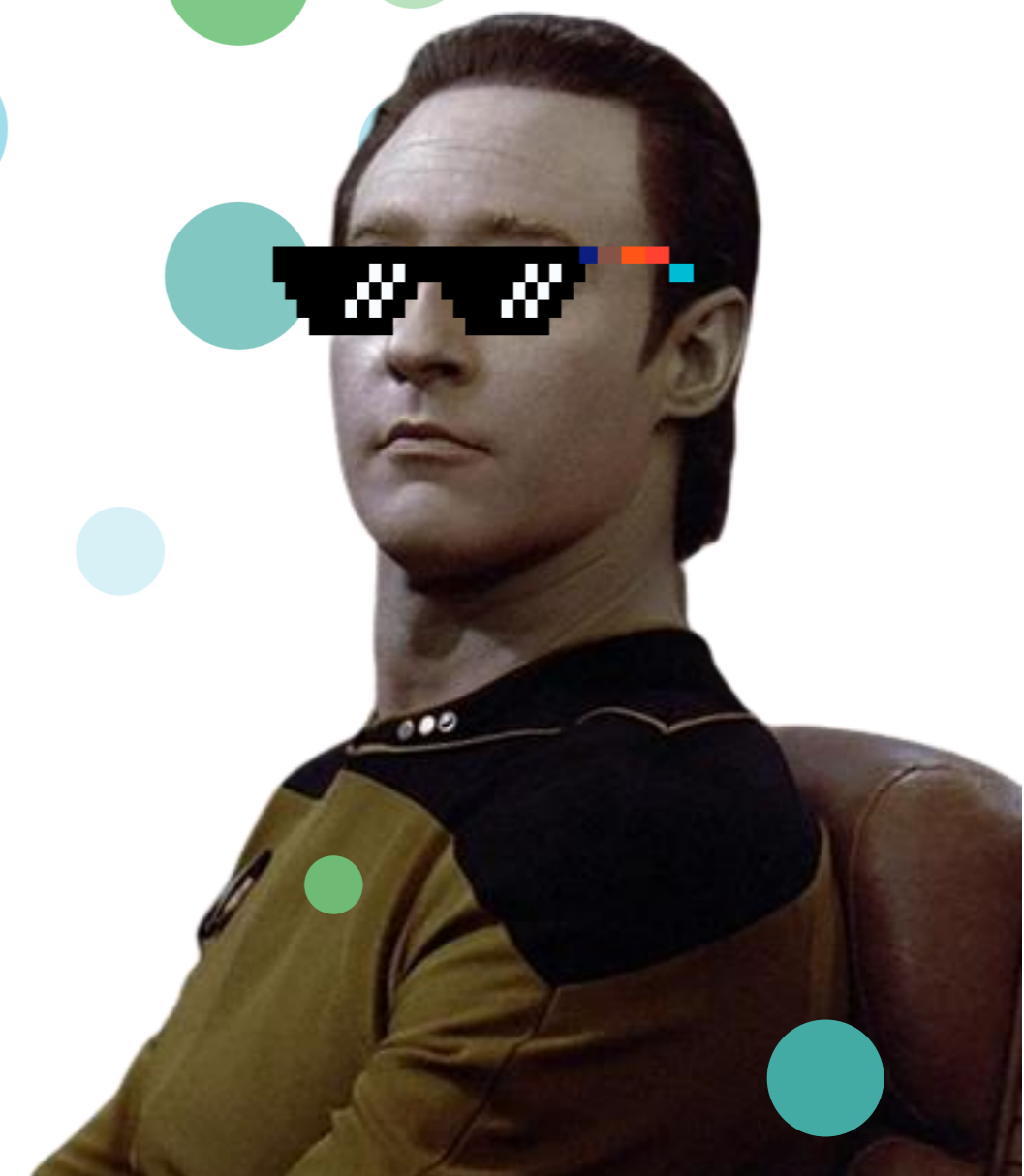


# Architecture: Rasa, SmartBot's Skeleton



# Knowledge Graph in Coreon: structuring Vienna Business Agency's Data

domain knowledge drafting



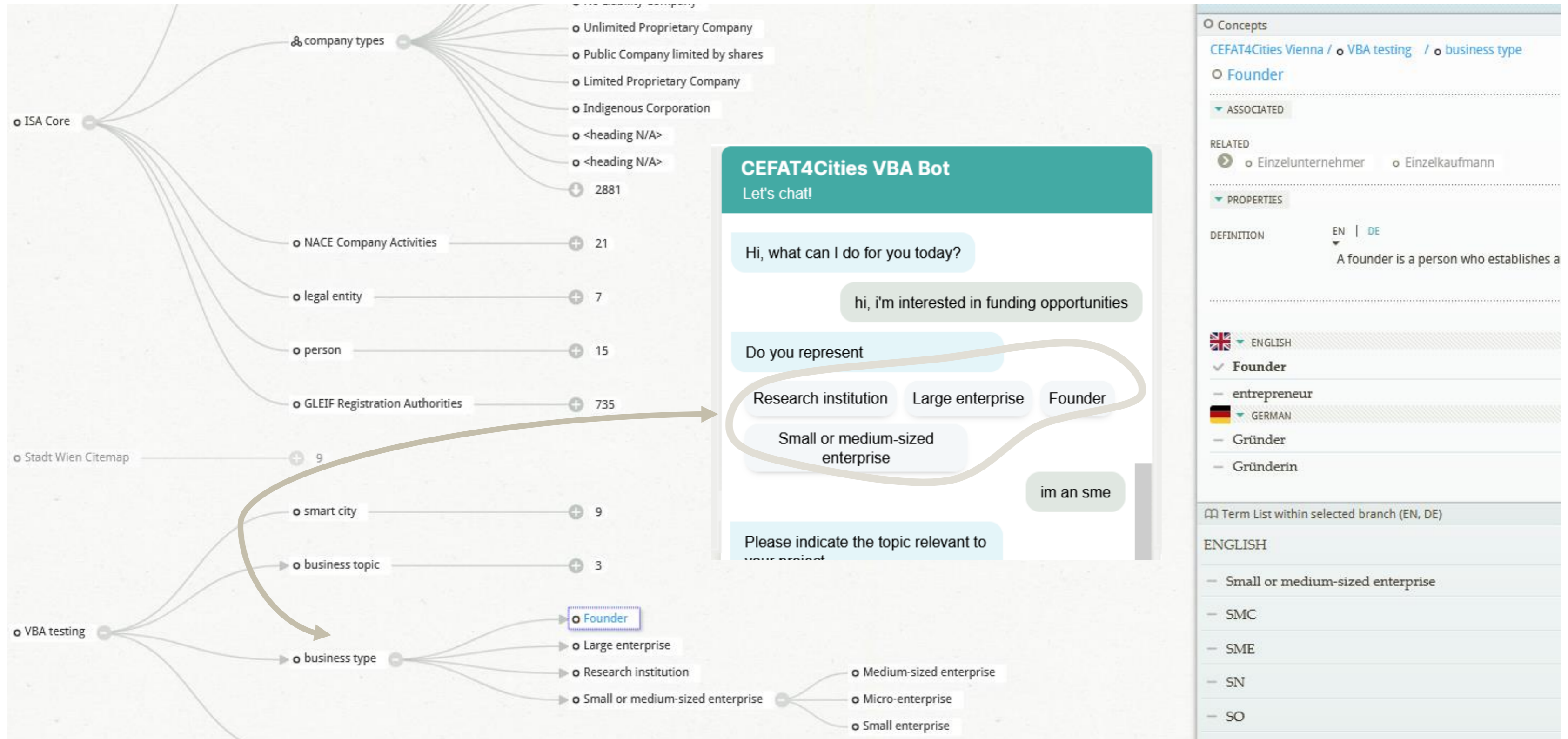


# Coreon: Easy Drafting of Knowledge for Domain Experts



- incorporate **taxonomies**  
**ontologies**  
**thesauri**  
**vocabularies**
- concept-oriented and language-agnostic data model
- intuitive, lightweight **UI**
- create relations between concepts
- all in one API-callable **Knowledge Graph**

# Embedding VBA Domain Knowledge in Coreon



**ISA Core**

- company types
  - Unlimited Proprietary Company
  - Public Company limited by shares
  - Limited Proprietary Company
  - Indigenous Corporation
  - <heading N/A>
  - <heading N/A>
  - 2881
- NACE Company Activities (+ 21)
- legal entity (+ 7)
- person (+ 15)
- GLEIF Registration Authorities (+ 735)

**VBA testing**

- business topic (+ 3)
- business type
  - Founder
  - Large enterprise
  - Research institution
  - Small or medium-sized enterprise
    - Medium-sized enterprise
    - Micro-enterprise
    - Small enterprise

**CEFAT4Cities VBA Bot**  
Let's chat!

Hi, what can I do for you today?

hi, i'm interested in funding opportunities

Do you represent

Research institution Large enterprise Founder

Small or medium-sized enterprise

im an sme

Please indicate the topic relevant to your project

**Coreon Ontology View**

CEFAT4Cities Vienna / VBA testing / business type

**Founder**

ASSOCIATED

RELATED

- Einzelunternehmer
- Einzelkaufmann

PROPERTIES

DEFINITION EN | DE

A founder is a person who establishes a

ENGLISH

- entrepreneur
- Gründer
- Gründerin

GERMAN

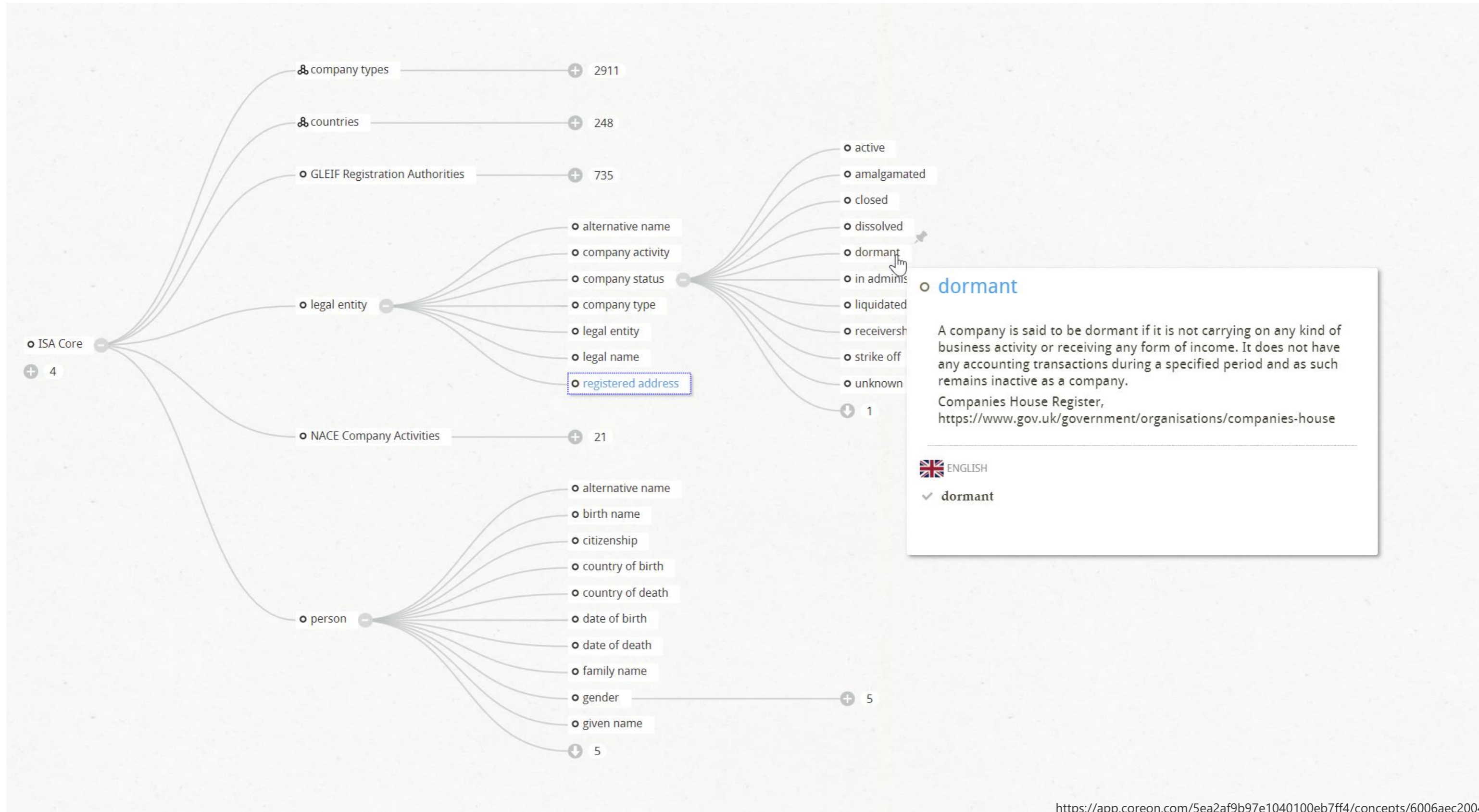
- Gründer
- Gründerin

Term List within selected branch (EN, DE)

ENGLISH

- Small or medium-sized enterprise
- SMC
- SME
- SN
- SO

# Vocabularies Incorporated into MKS



<https://app.coreon.com/5ea2af9b97e1040100eb7ff4/concepts/6006aec200408c0032ea2e41>

## Leveraging language-agnostic KGs:

- multilinguality
- entity management
- semantic search
- dealing with homonymy & unseen terms



## How to serve same info in different languages?

- Language detection vs manual selection in UI?
- Should Bot “understand” all languages but reply in one?
  - Or rather reply in the language of the last message?

## Our approach:

- NLU models per language
- One universal Core model/Stories

## Challenge:

Stories contain entities aka **language-specific** keywords  
Thus should be mapped to some language-independent IDs

# Rasa x Coreon: Entities & Their Management



```
1 {
2   "text": "hi, i'm interested in subsidies",
3   "intent": {
4     "id": 8462830655171089267,
5     "name": "want_funding",
6     "confidence": 0.5987483859
7   },
8   "entities": [
9     {
10      "entity": "funding",
11      "start": 22,
12      "end": 31,
13      "confidence_entity": 0.9987052679,
14      "value": "61adf0426951890d44fa2574",
15      "extractor": "DIETClassifier"
16    }
17  ],
18  "intent_ranking": [
19    {
20      "id": 8462830655171089267,
21      "name": "want_funding",
22      "confidence": 0.5987483859
23    }
24  ],
25 }
```

- abstract from entity maintenance in distinct languages
- replace language-specific terms with their unique Coreon concept IDs in the language-specific NLU training data
- this approach keeps data clean and consistent
- facilitates use of the one Core NLG model across languages
- if VBA decides to add a new language to SmartBot, "universal entities" will ensure consistency, smooth model development, and minimization of labeling effort

# Rasa x Coreon: Relations For Semantic Search



The screenshot displays a semantic search interface with a concept graph on the left and detailed information for a selected concept on the right.

**Concept Graph:**

- ISA Core
  - company types (2881)
    - Unlimited Proprietary Company
    - Public Company limited by shares
    - Limited Proprietary Company
    - Indigenous Corporation
    - <heading N/A>
    - <heading N/A>
  - NACE Company Activities (21)
  - legal entity (7)
  - person (15)
  - GLEIF Registration Authorities (735)
- Stadt Wien Citemap (9)
  - smart city (9)
  - business topic (3)
    - Founder
    - Large enterprise
    - Research institution
    - Small or medium-sized enterprise
      - Medium-sized enterprise
      - Micro-enterprise
      - Small enterprise
- VBA testing
  - business type

**Selected Concept: Founder**

**CEFAT4Cities Vienna / VBA testing / business type**

ID: 6128f59669518969177f2439  
 CREATED: 2021-08-27T14:24:22.384Z (Alena Vasilevich)  
 UPDATED: 2021-09-13T16:00:26.272Z (Alena Vasilevich)

**Founder**

**BROADER & NARROWER**

- business type { Founder { No further narrower concepts

**ASSOCIATED**

RELATED: Einzelunternehmer, Einzelkaufmann

**PROPERTIES**

DEFINITION: EN | DE  
 A founder is a person who establishes a...

**ENGLISH**

- entrepreneur
- Gründer
- Gründerin

**GERMAN**

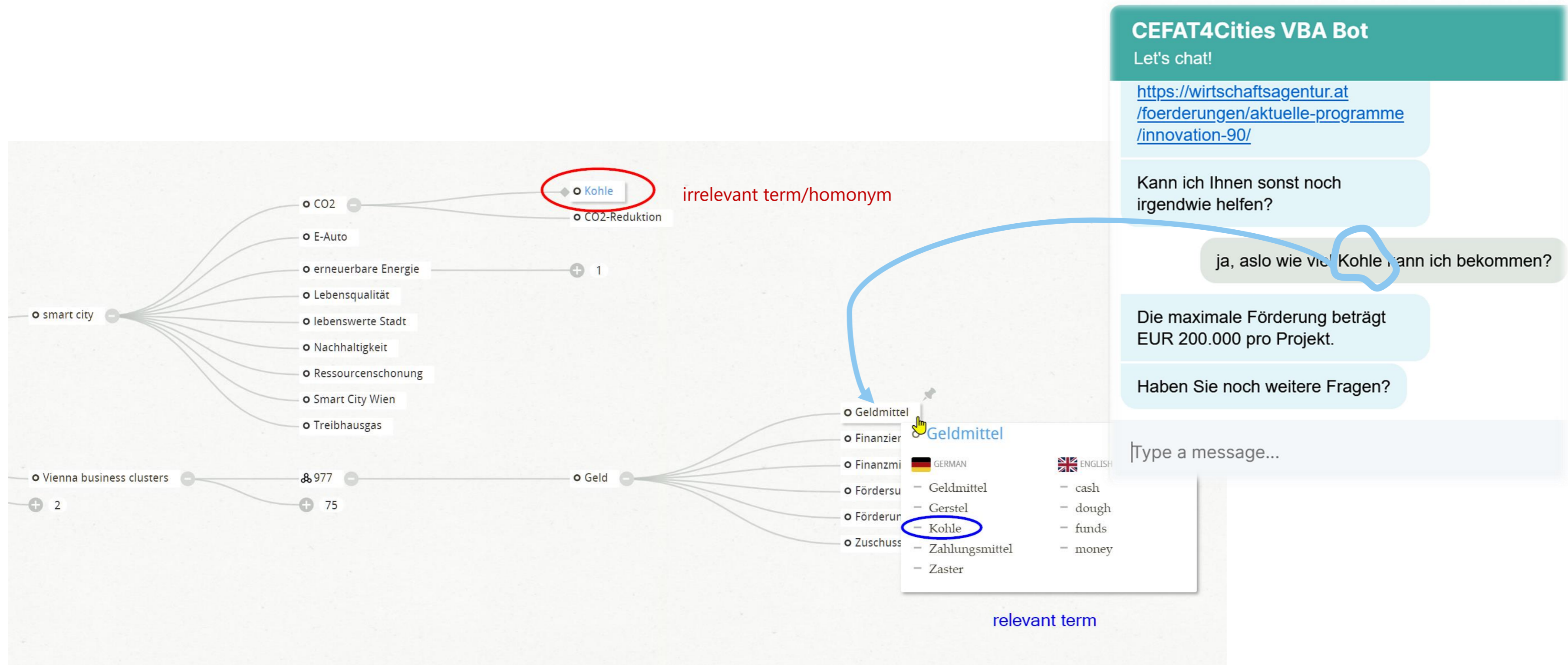
- entrepreneur
- Gründer
- Gründerin

**Term List within selected branch (EN, DE)**

**ENGLISH**

- Small or medium-sized enterprise
- SMC
- SME
- SN
- SO

# Rasa x Coreon: Homonymy and Terms Unseen by Model



**smart city**

- CO2
- E-Auto
- erneuerbare Energie
- Lebensqualität
- lebenswerte Stadt
- Nachhaltigkeit
- Ressourcenschonung
- Smart City Wien
- Treibhausgas

**Vienna business clusters**

- 977
- 75

**Geld**

- Geldmittel
- Finanzier
- Finanzmi
- Fördersu
- Förderun
- Zuschuss

**Kohle** (irrelevant term/homonym)

**Geldmittel** (relevant term)

- GERMAN
  - Geldmittel
  - Gerstel
  - Kohle
  - Zahlungsmittel
  - Zaster
- ENGLISH
  - cash
  - dough
  - funds
  - money

**CEFAT4Cities VBA Bot**

Let's chat!

<https://wirtschaftsagentur.at/foerderungen/aktuelle-programme/innovation-90/>

Kann ich Ihnen sonst noch irgendwie helfen?

ja, aslo wie viel Kohle kann ich bekommen?

Die maximale Förderung beträgt EUR 200.000 pro Projekt.

Haben Sie noch weitere Fragen?

Type a message...



# SmartBot Powered by Multilingual AI (Rasa x Coreon)

Technically robust

Intelligent chatbot solution, scalable, extendable, and modular

Resources reusability

VBA's proprietary KG + results of CEFAT4Cities action  
+ reuse of other EU data sources

Multilinguality

Consistency in multilingual data management;  
helps overcome the language gap in a user-friendly way

Red Tape Reduction

Rationalization of business processes, easy VBA service discoverability,  
elevation of local public services on the European scale



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

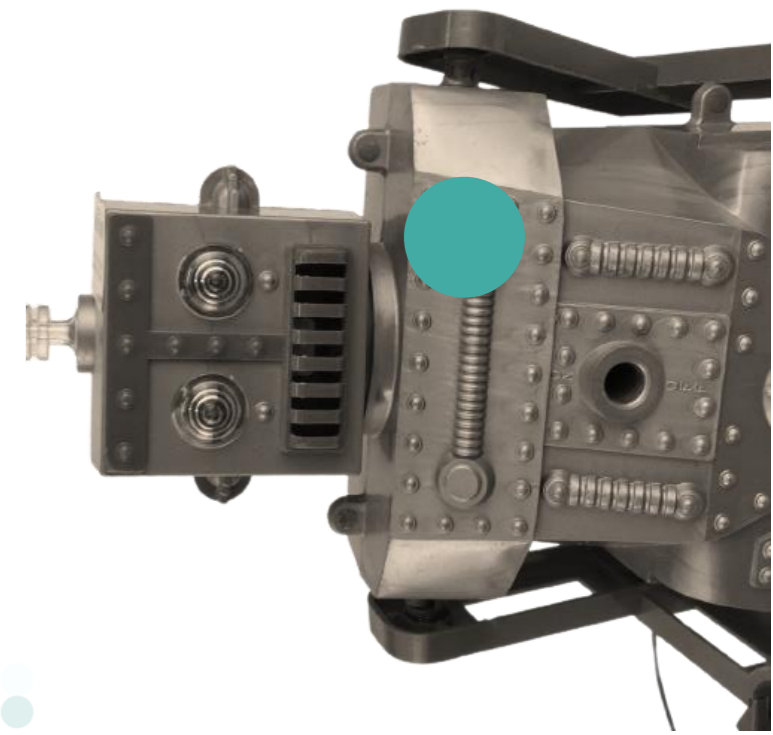


# coreon

Knowledge meets language.

**#KGC2022**

Thank you!



@coreonapp

@lennyvasilevich



<https://www.linkedin.com/company/coreon-gmbh>

<https://www.linkedin.com/in/alenasvasilevich>



The  
Knowledge  
Graph  
Conference