



PROJECT

AUTOMATIC CBO UPDATING



Supply Chain

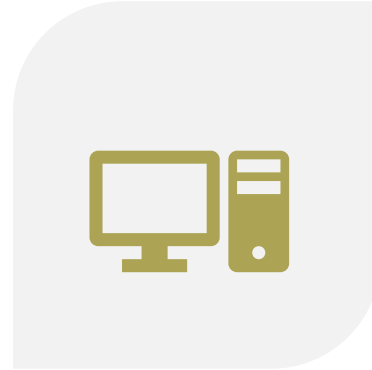


Ministério do Trabalho e Emprego
Subsecretaria de Estudos e Estatísticas do Trabalho

GUIDE



INTRODUCTION

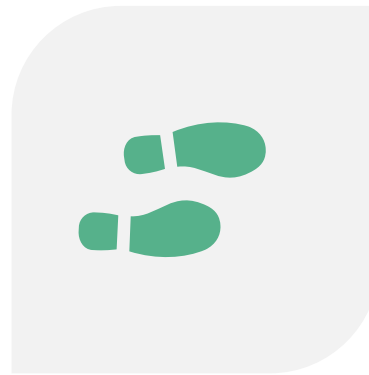


DEVELOPMENT



RESULTS

NEXT STEPS



INTRODUCTION

Motivation

- ✓ Apply artificial intelligence to help improve public services in the occupation classification field.
- ✓ Create a development process that could be replicated in other projects.

Old Upd Process

- ✓ Data bases Search done manually.
- ✓ Databases hosted in different places and systems.
- ✓ Lack of unified view of the process.
- ✓ Low response time.

INTRODUCTION

Objectives

- ✓ Create a model based in artificial intelligence (natural language processing) to automate steps of the CBO* updating process and assist the analyst.
- ✓ Create a flexible and replicable development process.

Challenges

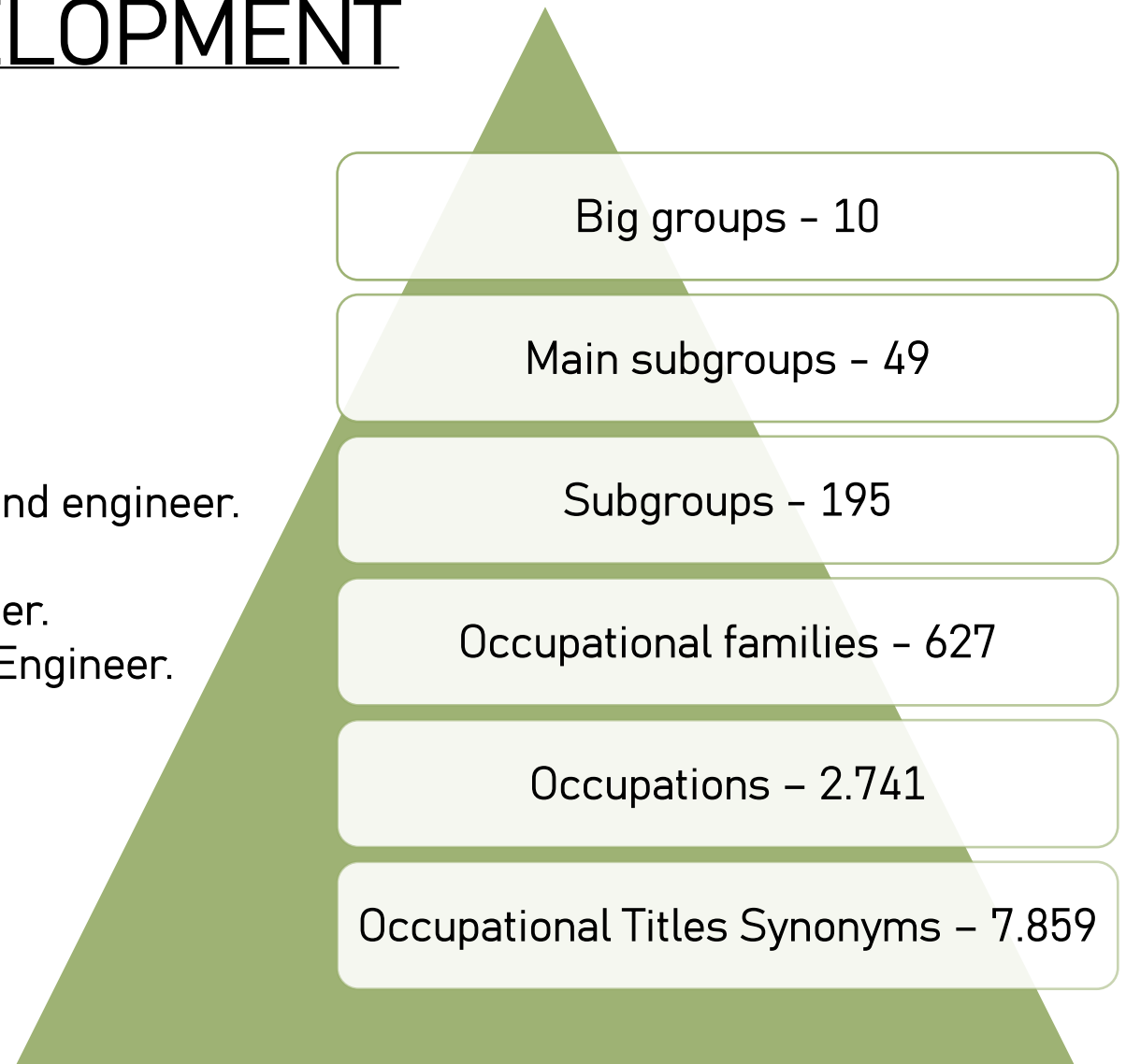
- ✓ Starting from scratch.
- ✓ Gather and prepare the data.
- ✓ Be incremental.
- ✓ Follow the technological changes in the labor market.
- ✓ Set up the right strategy.
- ✓ Set up the computation environment.
- ✓ Create a friendly front end.

* CBO – Classificação Brasileira de Ocupações (Brazilian Occupation Classification)

DEVELOPMENT

CBO Structure

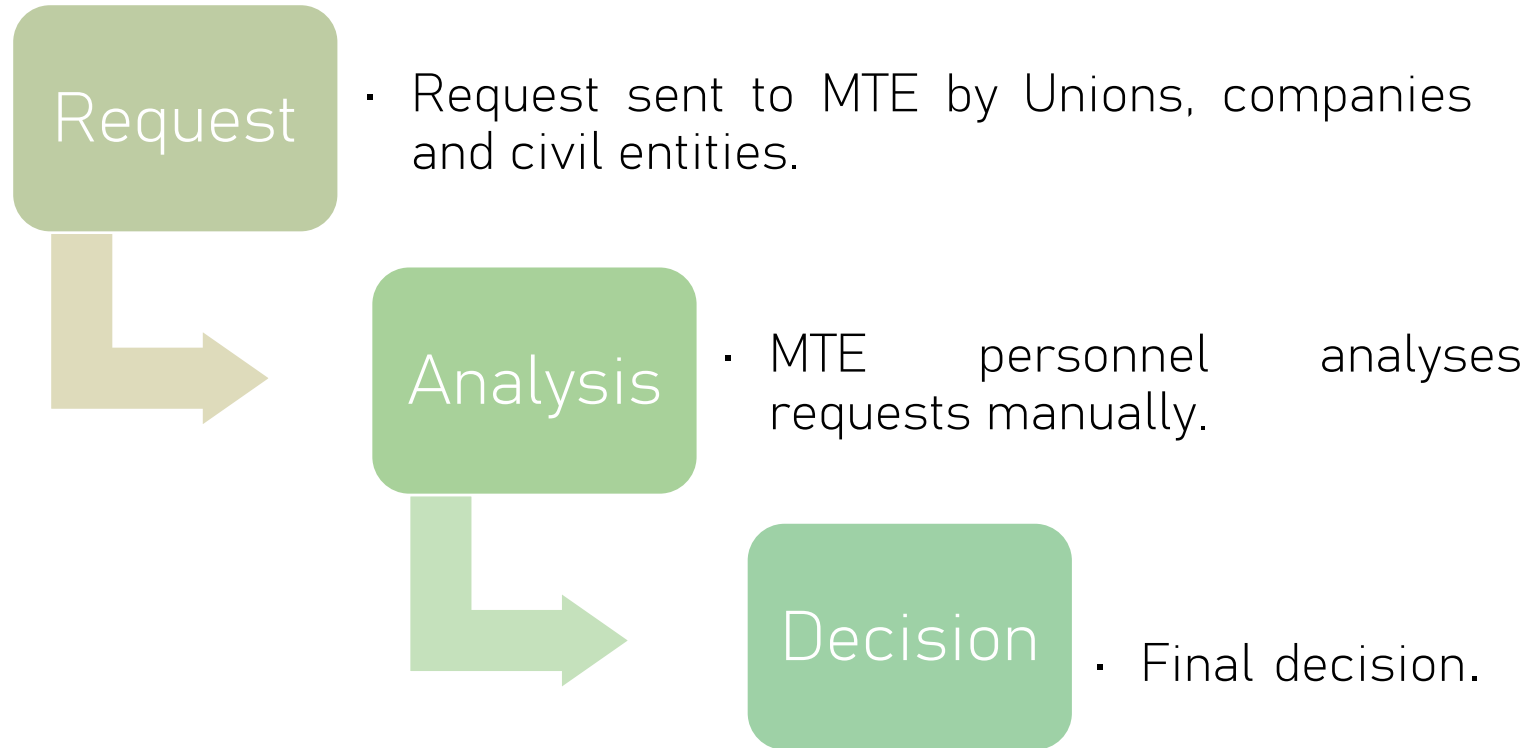
- ✓ Hierarquical structure.
- ✓ Example: Engineer.
- ✓ Big group: 2 – Science and Arts.
- ✓ Main subgroup: 21 – Exact science, physics and engineer.
- ✓ Subgroup: 214 – Engineer and Architect.
- ✓ Occupation family: 2144 – Mechanical Engineer.
- ✓ Occupation: 214410: Automotive Mechanical Engineer.
- ✓ Synonyms:
 1. Car maintenance engineer.
 2. Automobile mechanical engineer.
 3. Mechanical engineer (vehicle).



DEVELOPMENT

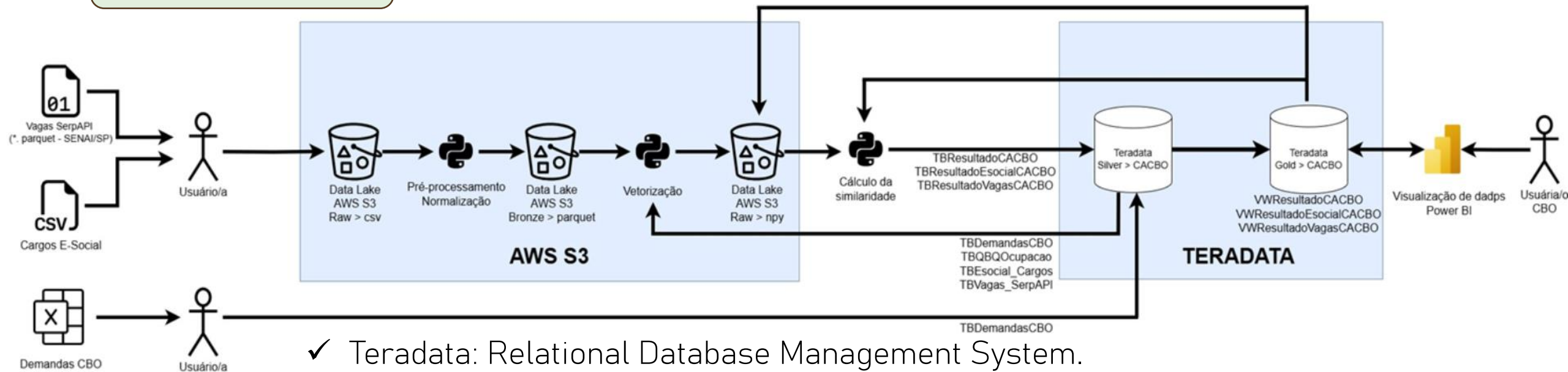
CBO Updating

✓ Old Process.



DEVELOPMENT

Infrastructure



- ✓ Teradata: Relational Database Management System.
- ✓ AWS3: Cloud-based object storage service.
- ✓ MTE Gitlab: Code Hosting.
- ✓ Python Server
- ✓ PowerBi: Front end

DEVELOPMENT

AI Models

- ✓ Language: Python.
- ✓ Data Preparation:
 - ✓ Normalization (stop words; tokenization, lemantization).
 - ✓ Vetorization (sparse, hugging face (dense)).
- ✓ Natural Language Processing (NLP) – FAISS: Facebook AI Similarity Search:
 - ✓ QBQ* data base: hugging face cosine.
 - ✓ Professional Profile + Knowlegde x Demand occupation title.
 - ✓ eSocial** data base: hugging face cosine (22 million records).
 - ✓ Job Title x Demand occupation title.
 - ✓ **Job advertisements** data base: hugging face cosine.
 - ✓ Job Title x Demand occupation title (1.743.000 mil records).

*QBQ – Quadro Brasileiro de Qualificações (Brazilian Skillset List)

**eSocial – Electronic administrative labor records.

RESULTS

POWERBI

Explaining the results:

- ✓ Finding where the demand is located in the CBO structure.
- ✓ Finding the occupational families Frequency.
- ✓ Understanding if the demand is already considered in the CBO list or if It is a new occupation.

CBO Structure

Big groups - 10

Main subgroups - 49

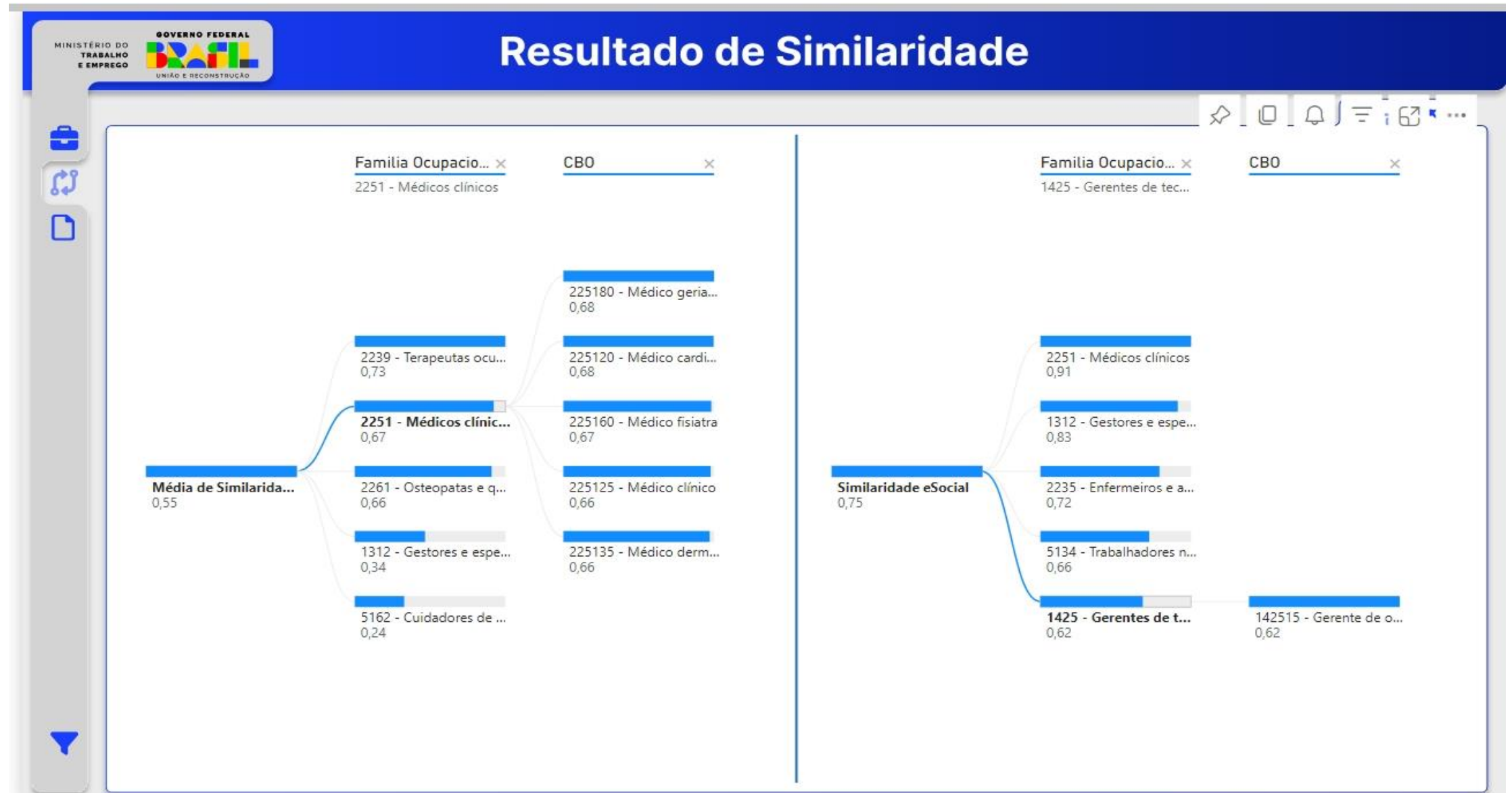
Subgroups - 195

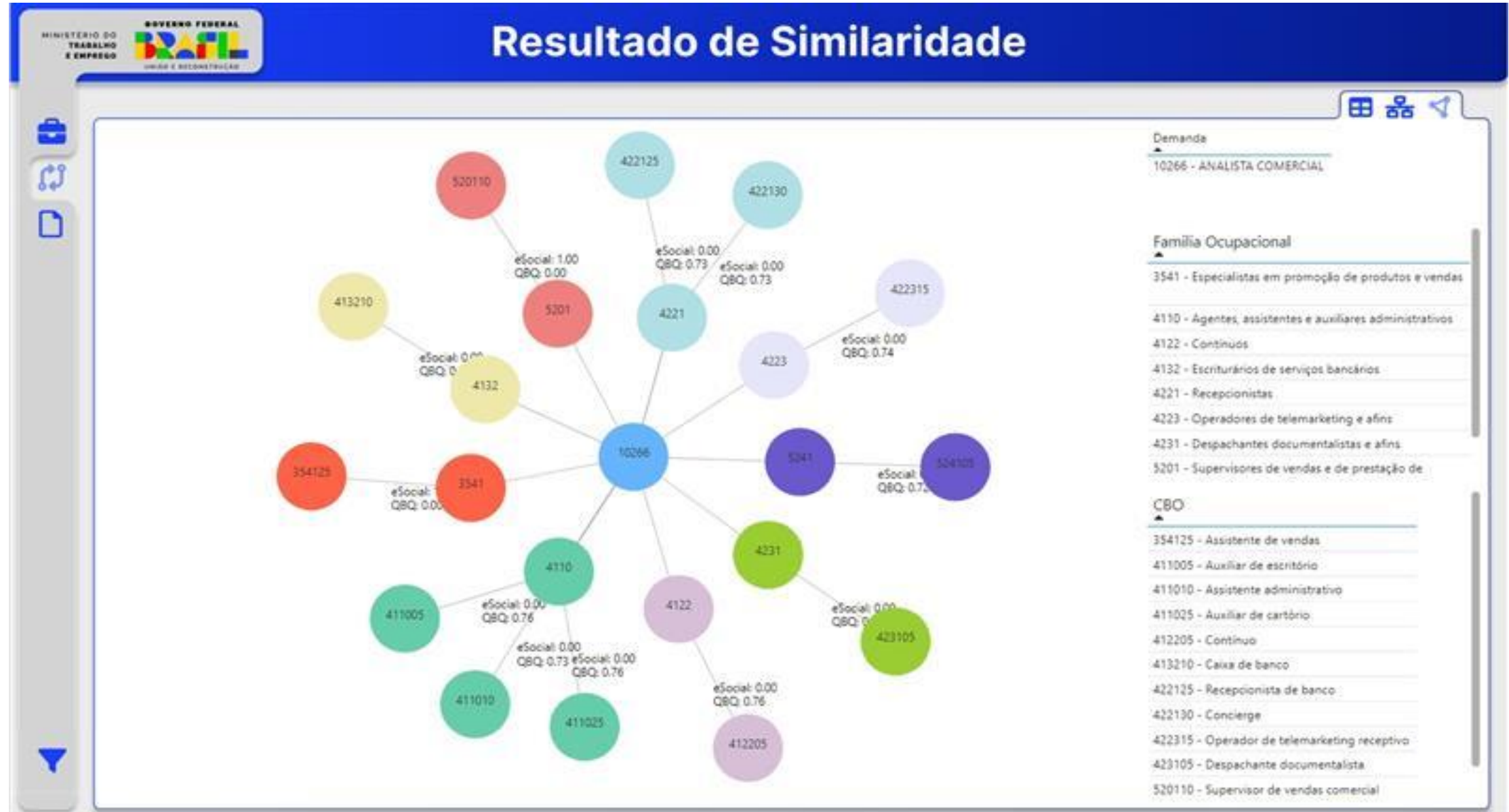
Occupational families - 627

Occupations - 2.741

Occupational Titles Synonyms - 7.859

Resultado de Similaridade					
GOVERNO FEDERAL MINISTÉRIO DO TRABALHO E EMPREGO UNIÃO E RECONSTRUÇÃO					
Similaridade QBQ	Similaridade eSocial	Similaridade Vagas			
Demanda	Família Ocupacional	CBO	Cargo eSocial	Similaridade eSocial	
10503 - Gerontóloga	1312 - Gestores e especialistas de operações em empresas, secretarias e unidades de serviços de saúde	131220 - Gerontólogo	GERONTOLOGA EDUCADORA	0,81	
10503 - Gerontóloga	1312 - Gestores e especialistas de operações em empresas, secretarias e unidades de serviços de saúde	131220 - Gerontólogo	GERONTOLOGO	0,89	
10503 - Gerontóloga	1312 - Gestores e especialistas de operações em empresas, secretarias e unidades de serviços de saúde	131220 - Gerontólogo	GERONTOLOGO PL	0,84	
10503 - Gerontóloga	1312 - Gestores e especialistas de operações em empresas, secretarias e unidades de serviços de saúde	131220 - Gerontólogo	GF - GERONTOLOGO	0,79	
10503 - Gerontóloga	1425 - Gerentes de tecnologia da informação	142515 - Gerente de operação de tecnologia da informação	Gerente de Tecnologia e Gesto	0,62	
10503 - Gerontóloga	2235 - Enfermeiros e afins	223505 - Enfermeiro	ENFERMEIRO GERONTOLOGO	0,72	
10503 - Gerontóloga	2251 - Médicos clínicos	225180 - Médico geriatra	GERONTOLOGA	1,00	
10503 - Gerontóloga	2251 - Médicos clínicos	225180 - Médico geriatra	GERONTOLOGO	0,89	
10503 - Gerontóloga	2251 - Médicos clínicos	225180 - Médico geriatra	Gerontologo I	0,84	
10503 - Gerontóloga	5134 - Trabalhadores no atendimento em estabelecimentos de serviços de alimentação, bebidas e hotelaria	513405 - Garçom	GARON EM GERAL	0,66	

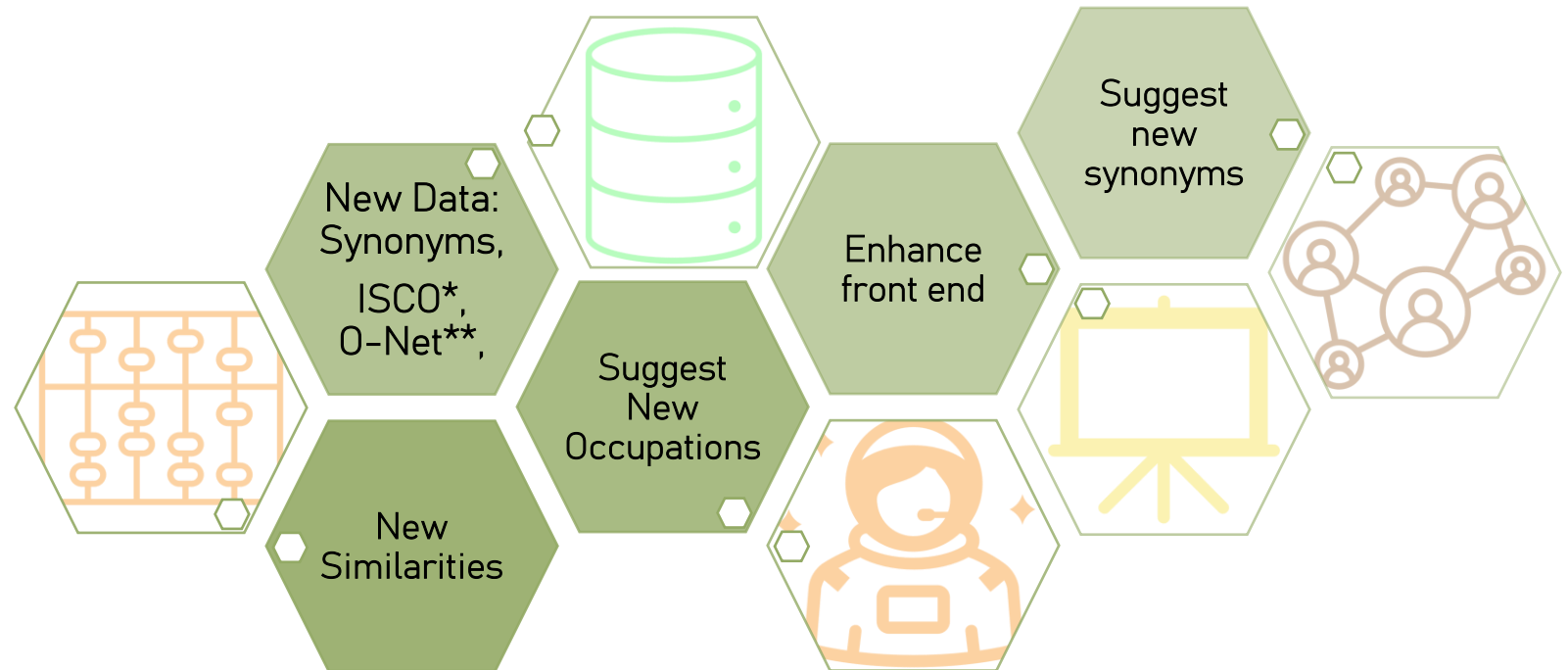




NEXT STEPS

GOALS

- ✓ Keep CBO up to date.
- ✓ Keep Synonyms up to date.
- ✓ Make data easily available.



*ISCO – International Standard Classification of Occupation.

**O-NET – Occupational Information Network.

2 DECEMBER 2024

BRASIL PROJECT AUTOMATIC CBO UPDATING

Thank you.

Team:

Name	Position
Paula Montagner:	Undersecretary
Enádio Barbosa	Project Manager
Claudio Bonel	AI developer
Aline Martins	CBO Head Team
Gabriela Nantes	CBO team member
Thomas Gomes	PBI Designer

Links:

Name	Link
Brazilian Occupation Classification:	http://www.mtecbo.gov.br/cbosite/pages/home.jsf
Brazilian Occupation Guide	https://www.gov.br/trabalho-e-emprego/pt-br/assuntos/estatisticas-trabalho/guia-brasileiro-de-ocupacoes