MEET EIFFEL: ROLE, CHALLENGES, EXPECTATIONS

EIRINI NTAROUIS

OUTLINE

- EIFFEL Overview
- LIBRA’s Role
- Technical/NLP Introduction
- LIBRA’S Technical Tasks
- Challenges
- Impact

EIFFEL Overview

LIBRA's Role

Technical/NLP Introduction

LIBRA'S Technical Tasks

Challenges

Impact

PROJECT INTRODUCTION

- **Title**: Revealing the role of **GEOSS** as the default digital portal for building ClimateChange adaptation & mitigation applications
- **Duration**: 36 months
- **Participants**: 19 partners; 8 countries

Eiffel

GEOSS APPLICATIONS
FOR CLIMATE CHANGE

GEOSS

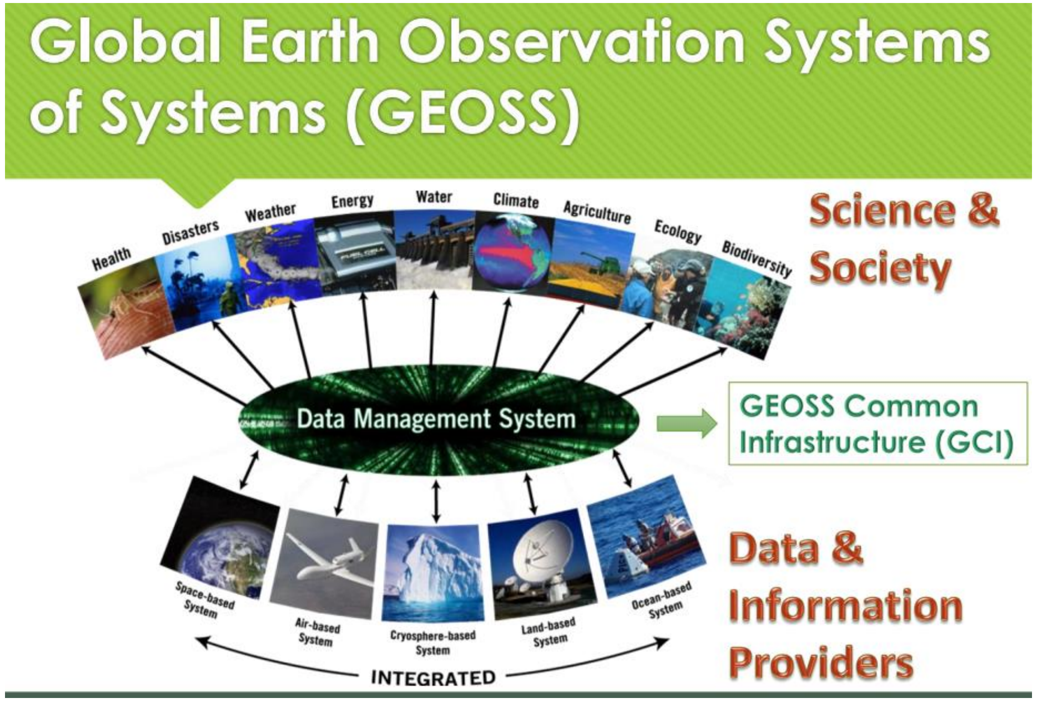
- What is GEOSS?

Global Earth Observation System of Systems is a set of coordinated, independent Earth observation systems.

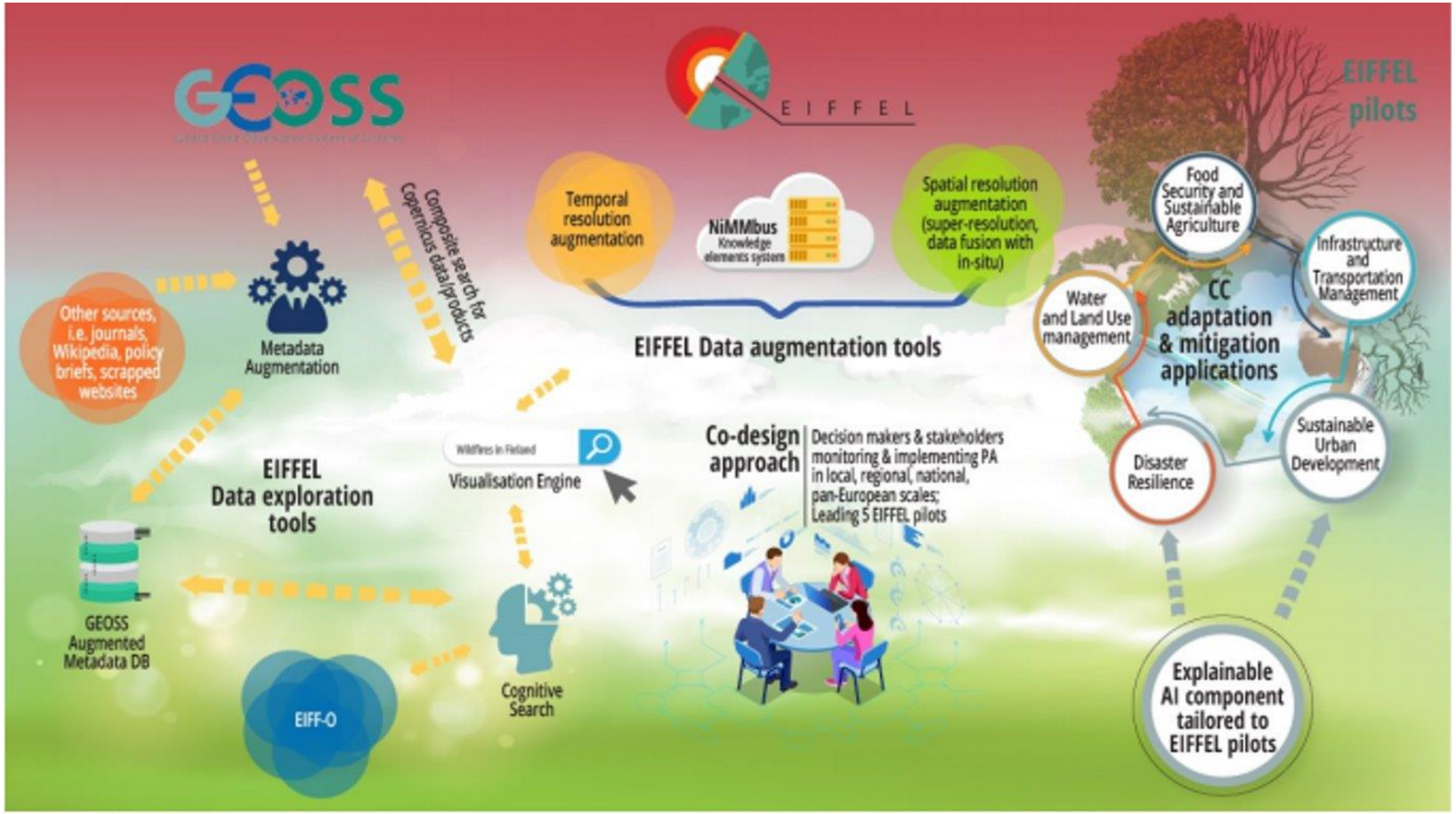
- What is Earth Observation?

The gathering of information about planet Earth's physical, chemical and biological systems.

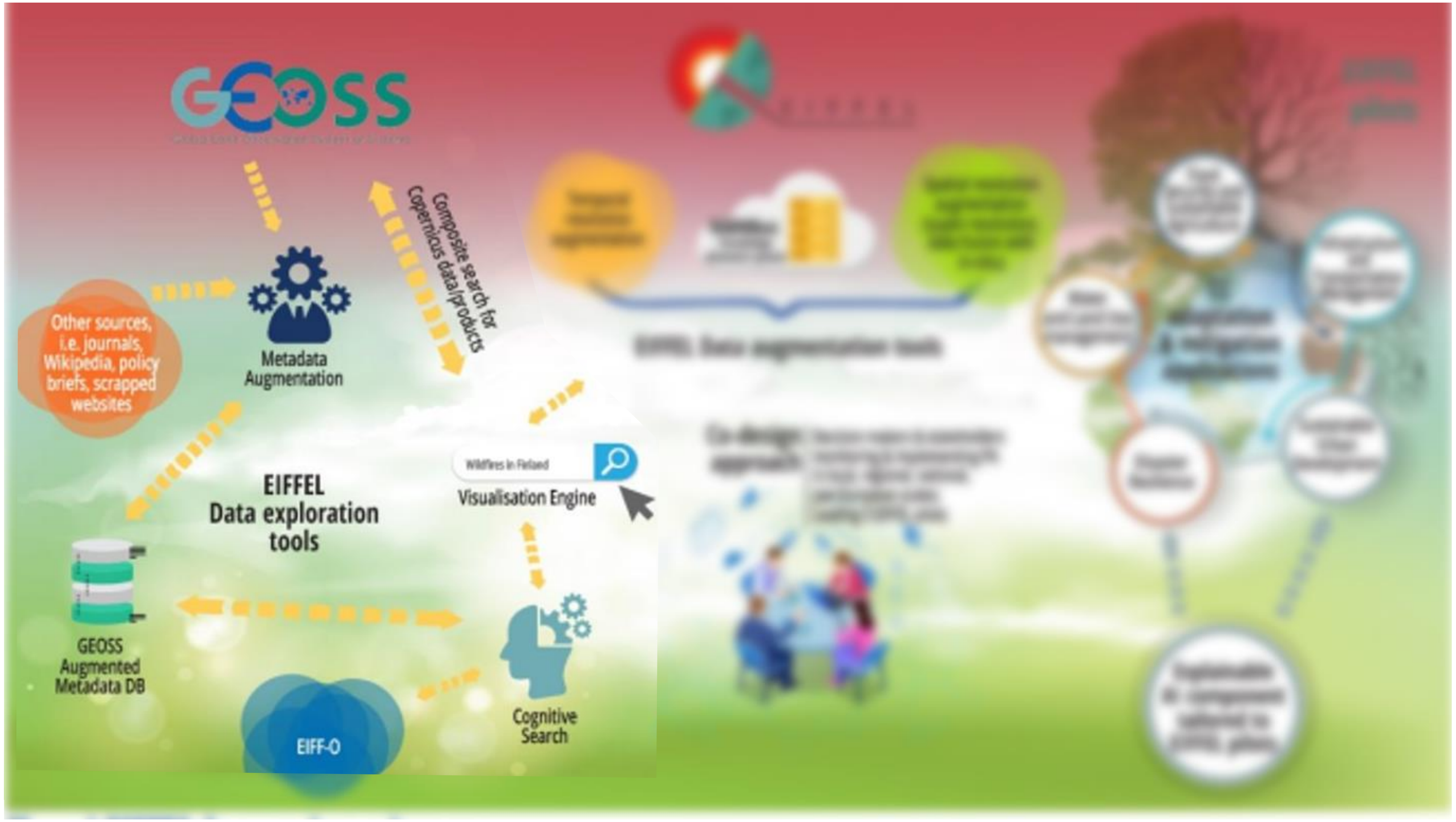
- Why is Earth Observation important?
  - is used to monitor and assess the status of, and changes in, the natural and manmade environment
  - **invaluable for assessing and mitigating the negative impacts of CC**
  - use for exploiting new opportunities, such as the sustainable management of natural resources



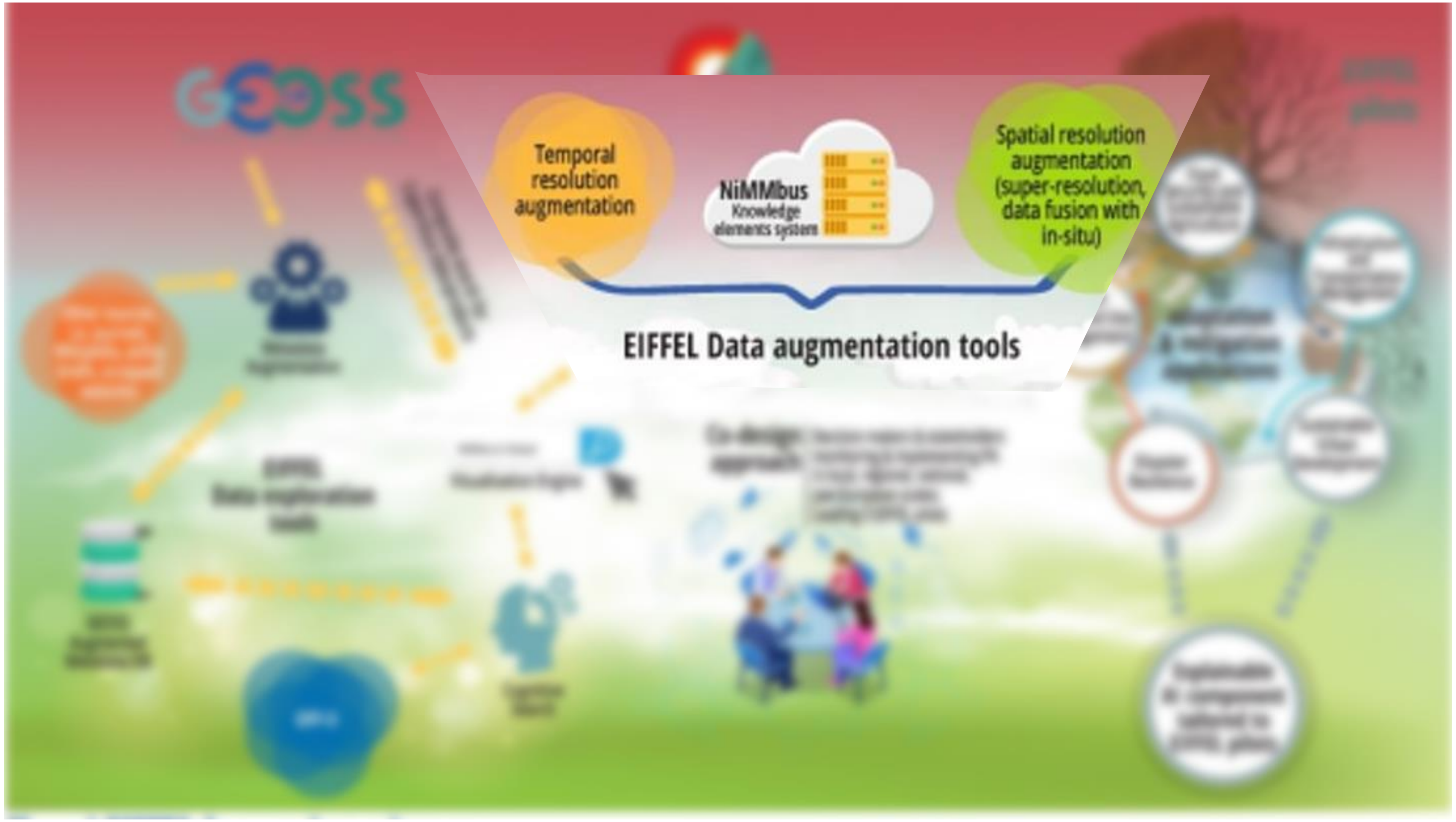
PROJECT OVERVIEW


GEOSS
EIFFEL
EIFFEL pilots
Other sources, i.e. journals, Wikipedia, policy briefs, scrapped websites
Metadata Augmentation
Composite search for Copernicus data/products
Temporal resolution augmentation
NiMMbus Knowledge elements system
Spatial resolution augmentation (super-resolution, data fusion with in-situ)
EIFFEL Data augmentation tools
Food Security and Sustainable Agriculture
Infrastructure and Transportation Management
Water and Land Use management
CC adaptation & mitigation applications
Sustainable Urban Development
Disaster Resilience
EIFFEL Data exploration tools
Wildfires in Finland
Visualisation Engine
GEOSS Augmented Metadata DB
Cognitive Search
EIFF-O
Co-design approach
Decision makers & stakeholders monitoring & implementing PA in local, regional, national, pan-European scales; Leading 5 EIFFEL pilots
Explainable AI component tailored to EIFFEL pilots

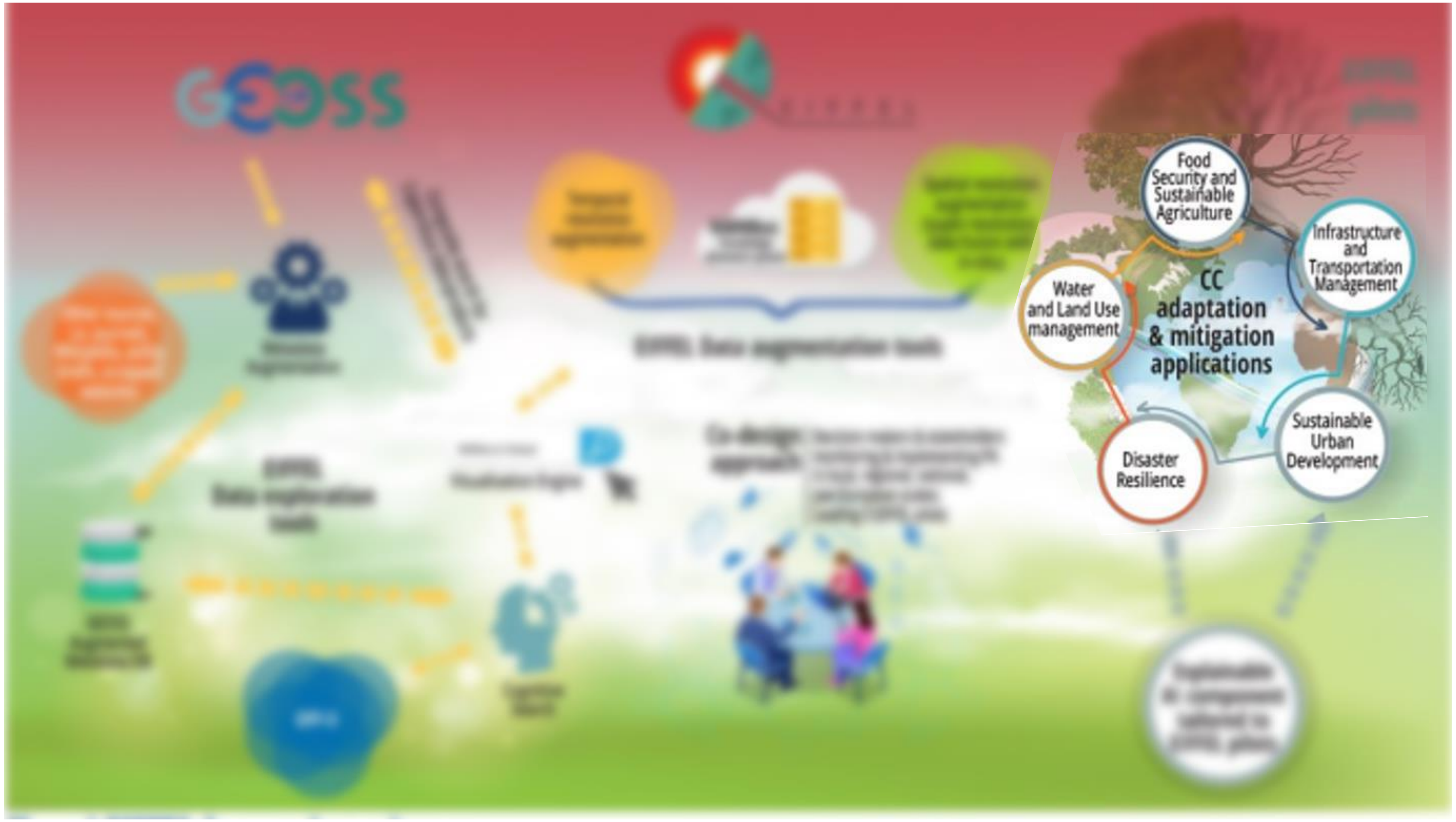
PROJECT OVERVIEW


GEOSS
Composite search for Copernicus data/products
Other sources, i.e. journals, Wikipedia, policy briefs, scrapped websites
Metadata Augmentation
EIFFEL Data exploration tools
Wildfires in Finland
Visualisation Engine
GEOSS Augmented Metadata DB
Cognitive Search
EIFF-O

PROJECT OVERVIEW


GEOSS
Temporal resolution augmentation
NiMMbus
Knowledge elements system
Spatial resolution augmentation (super-resolution, data fusion with in-situ)
EIFFEL Data augmentation tools

PROJECT OVERVIEW


GEOSS
Food Security and Sustainable Agriculture
Infrastructure and Transportation Management
Water and Land Use management
CC adaptation & mitigation applications
Sustainable Urban Development
Disaster Resilience

OUTLINE

EIFFEL Overview

LIBRA’s Role

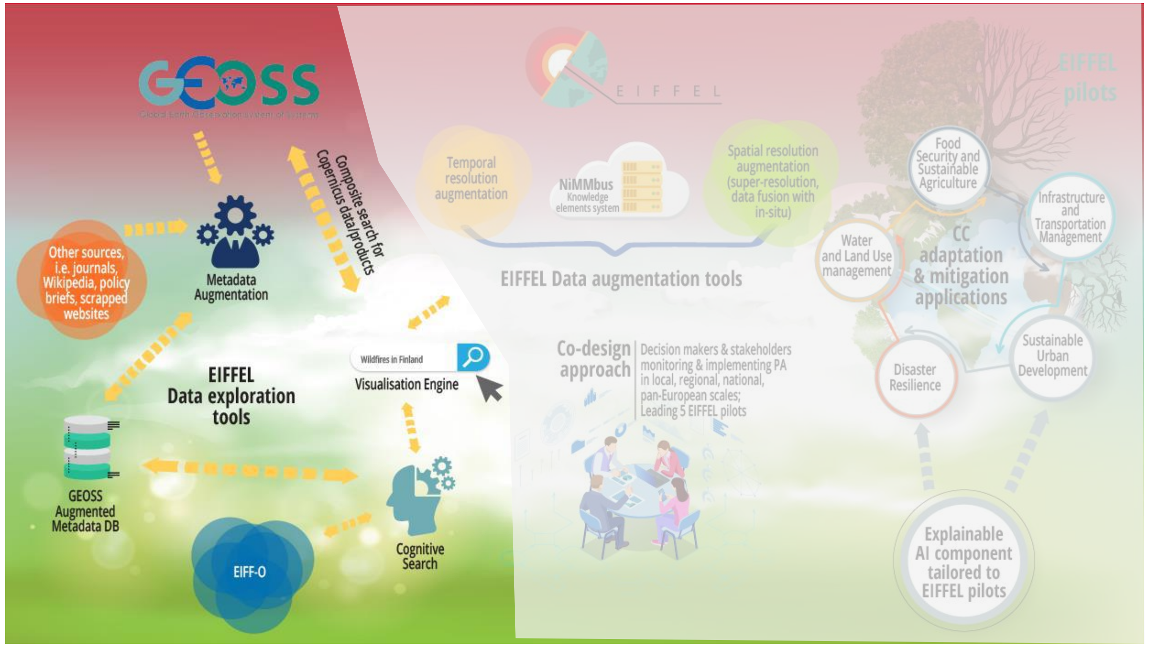
Technical/NLP Introduction

LIBRA’S Technical Tasks

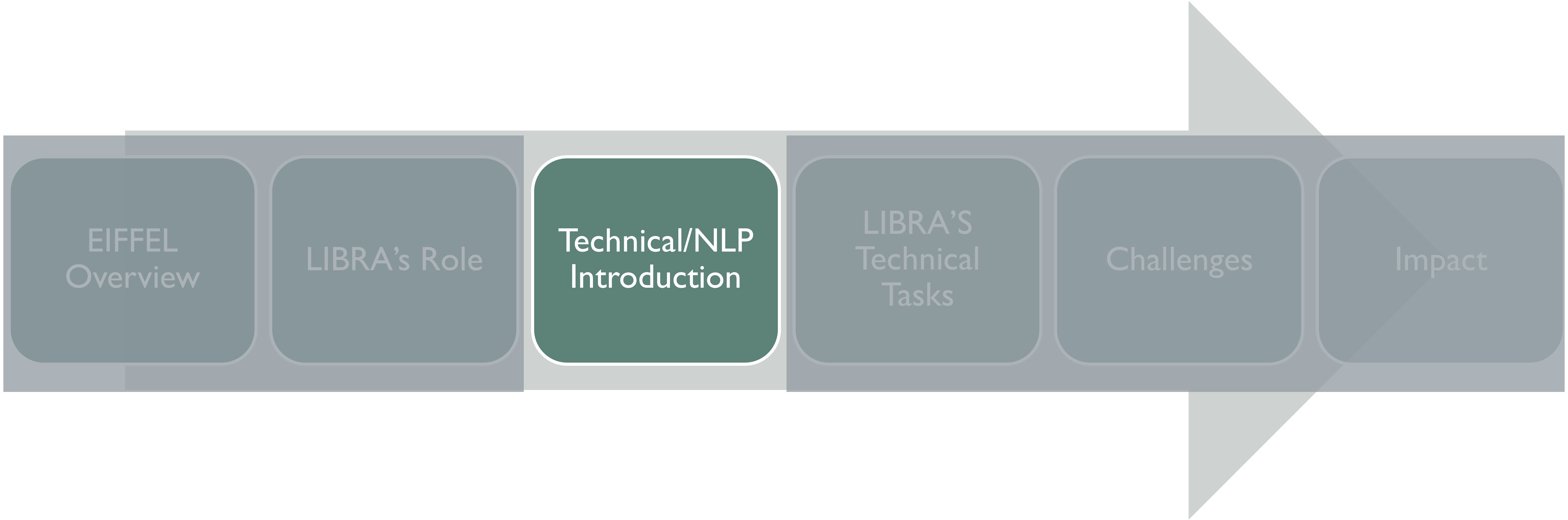
Challenges

Impact

LIBRA'S ROLE


GEOSS
Global Earth Observation System of Systems
Composite search for Copernicus data/products
Other sources, i.e. journals, Wikipedia, policy briefs, scrapped websites
Metadata Augmentation
EIFFEL Data exploration tools
Wildfires in Finland
Visualisation Engine
GEOSS Augmented Metadata DB
EIFF-O
Cognitive Search
EIFFEL
Temporal resolution augmentation
NiMMbus
Knowledge elements system
Spatial resolution augmentation (super-resolution, data fusion with in-situ)
EIFFEL Data augmentation tools
Co-design approach
Decision makers & stakeholders monitoring & implementing PA in local, regional, national, pan-European scales; Leading 5 EIFFEL pilots
EIFFEL pilots
Food Security and Sustainable Agriculture
Infrastructure and Transportation Management
Water and Land Use management
CC adaptation & mitigation applications
Disaster Resilience
Sustainable Urban Development
Explainable AI component tailored to EIFFEL pilots

- Leading the Augmenting GEOSS data exploration activities of the project, including:
  - ❖ NLP-based cognitive search engine
  - ❖ Visualisation engine of the EIFFEL cognitive search tool
  - ❖ Metadata curation and augmentation
  - ❖ EIFFEL CC-focused ontology
- Development of the cognitive search framework
- Design and deployment of metadata enrichment mechanisms



EIFFEL Overview

LIBRA's Role

Technical/NLP Introduction

LIBRA'S Technical Tasks

Challenges

Impact

- What is NLP?

  Machine learning applied to text / speech

- Computers only understand numbers, not characters, words, or sentences -> **text representation**

TRADITIONAL CONTEXT-FREE REPRESENTATIONS

- **Bag of Words**

  each element in the vector corresponds to a unique word in the vocabulary

  if word exists in document, element is marked as 1, else as 0

- **TF-IDF**

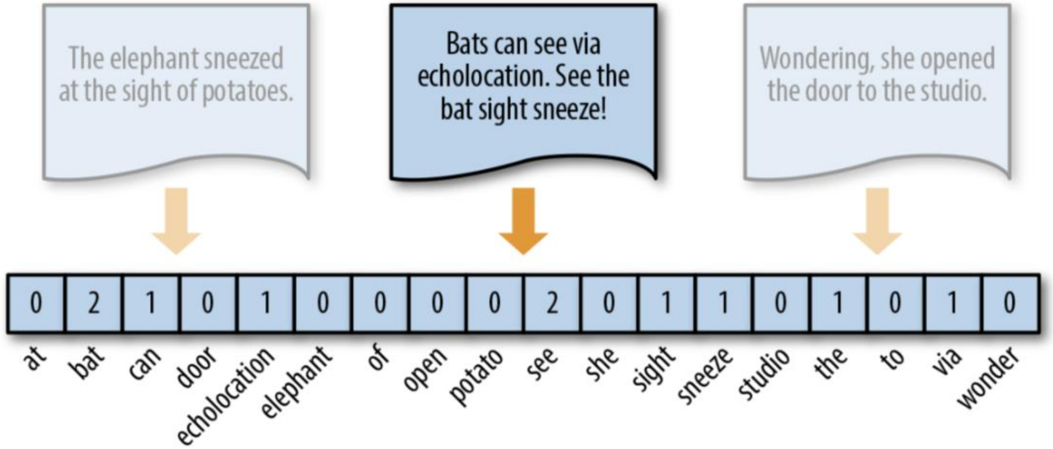
  TF: scoring of the frequency of the word in the current document

  IDF: scoring of how rare the word is across documents

  TF-IDF score = TF * IDF

  Contains information on the more/less important words

  UNABLE to capture word meaning/word similarity!



The elephant sneezed at the sight of potatoes.

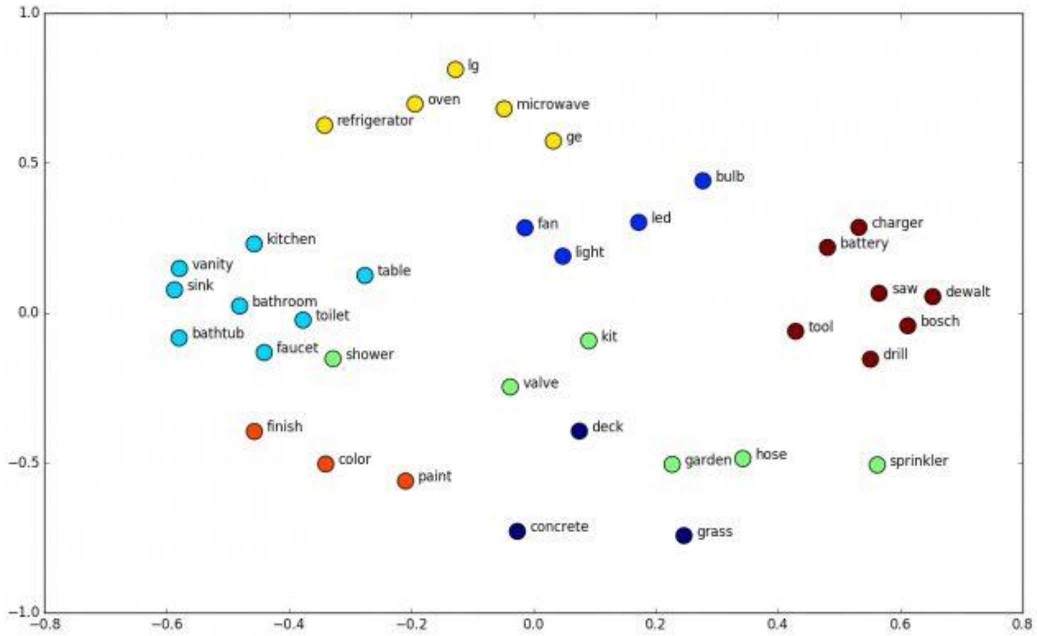
Bats can see via echolocation. See the bat sight sneeze!

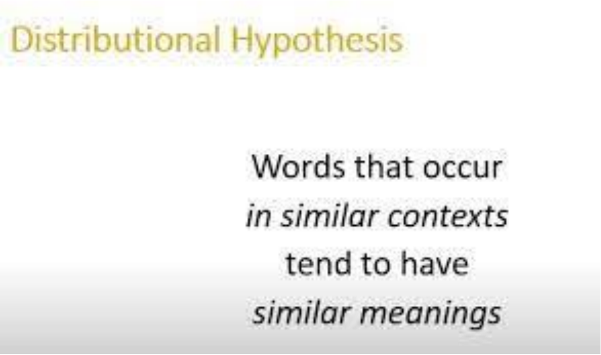
Wondering, she opened the door to the studio.

at	bat	can	door	echolocation	elephant	of	open	potato	see	she	sight	sneeze	studio	the	to	via	wonder
0	2	1	0	1	0	0	0	0	2	0	1	1	0	1	0	1	0

WORD EMBEDDINGS – DISTRIBUTIONAL SIMILARITY BASED REPRESENTATIONS

- Numeric vector representations of a particular word, that encodes the meaning of the word

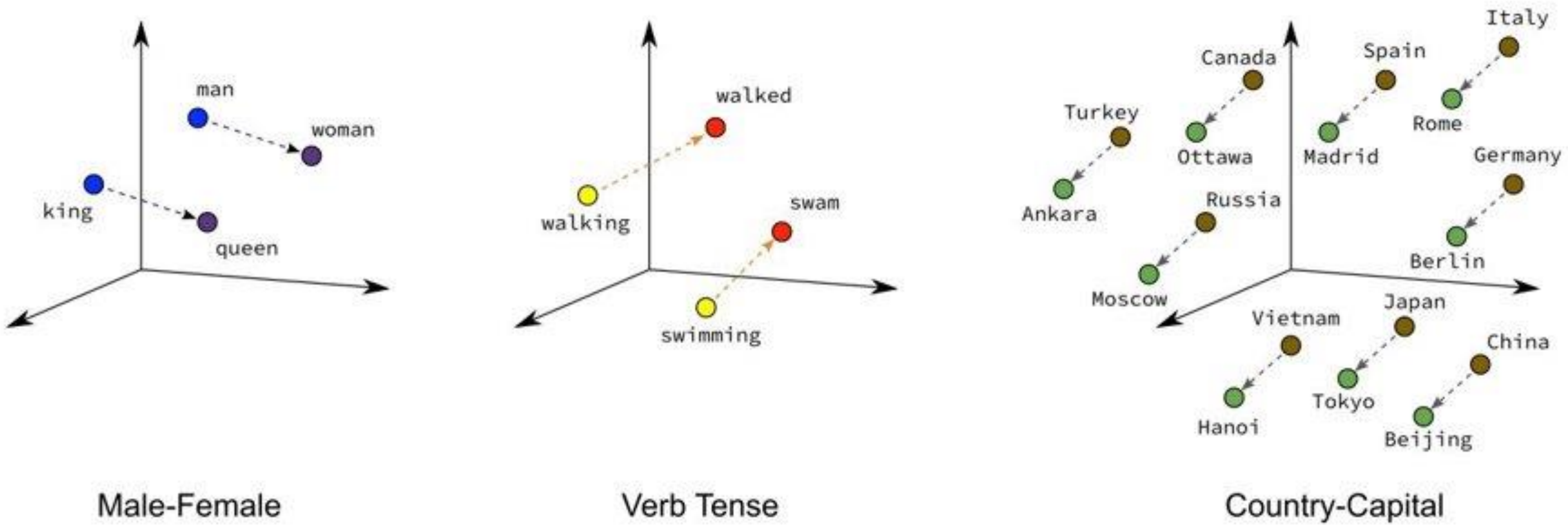



Distributional Hypothesis

Words that occur
in similar contexts
tend to have
similar meanings

WORD ANALOGIES

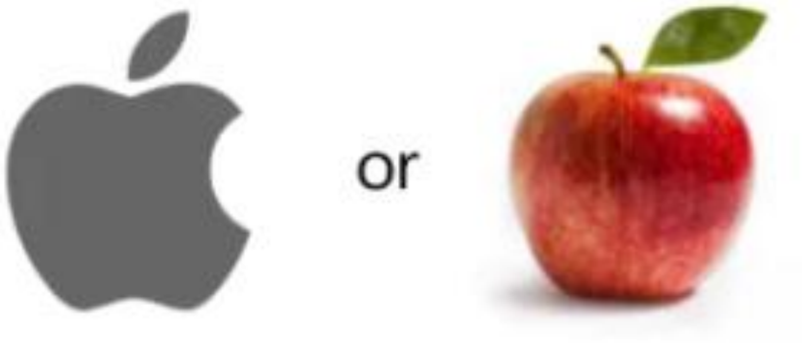
- A fascinating property of trained word embeddings is that the relationship between words is captured through linear relationships between vectors



king – man + woman = queen

LIMITATIONS OF WORD EMBEDDINGS

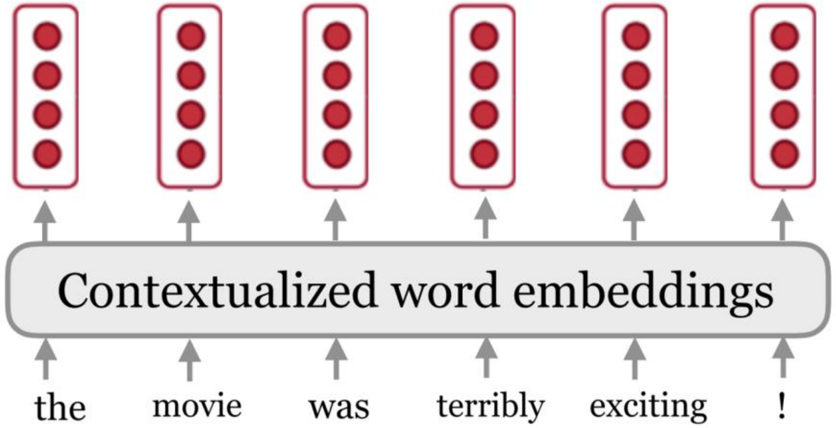
- One vector for each word type (static)
- Polysemous words, eg. Bank, mouse
- Words don’t appear in isolation. The word use depends on its context.
- Why not learn the representations for each word in its context?



“ I like apples” VS “I like Apple macbooks”

CONTEXTUALIZED WORD EMBEDDINGS

- Address the issue of polysemous and the context-dependent nature of words
- Build a vector for each word conditioned on its context!



$$g : (w_1, w_2, \ldots, w_n) \longrightarrow \mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$$

- Contextual representations of language leverage the intuition that the meaning of a particular word in a particular text depends not only on the identity of word, but also on the words that surround it at that moment

- ELMO, GPT-3 and BERT
- state-of-the-art performance on many popular tasks

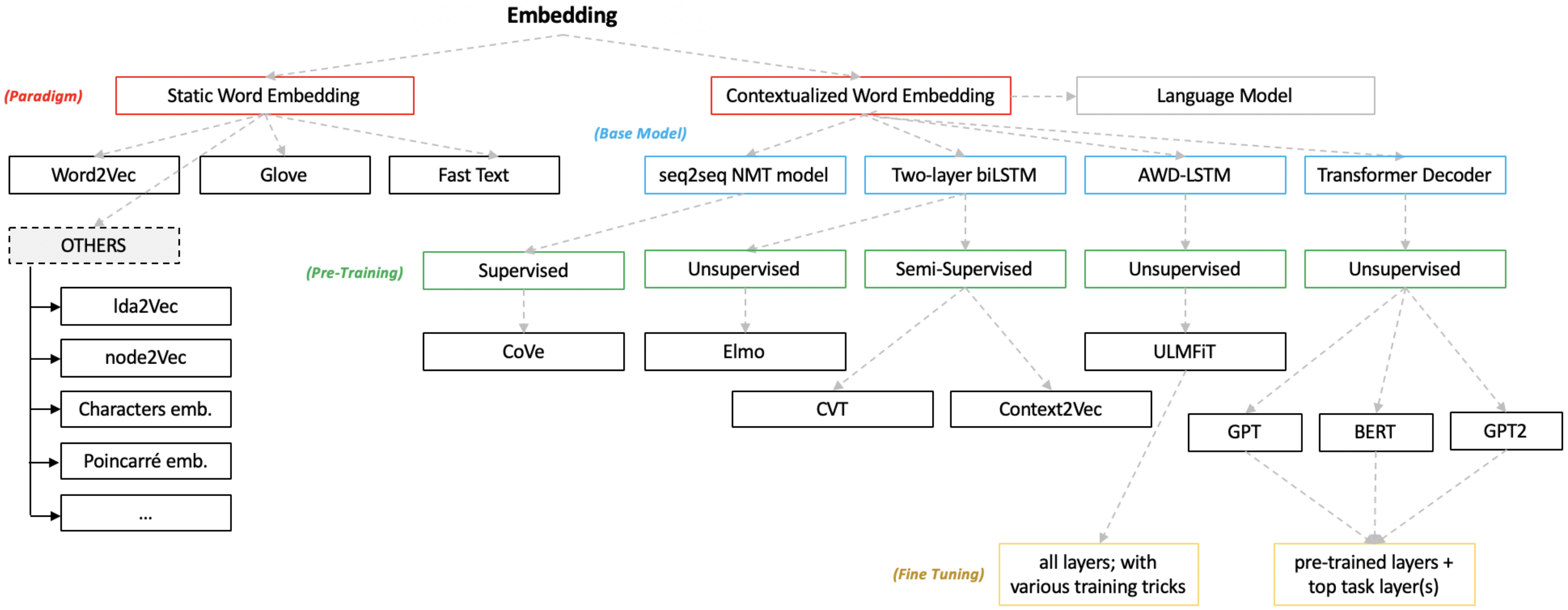
Take into account the word context

Trained on a large text corpus

Open-source

- Models pre-trained on language modeling (**unsupervised task**) and fine-tuned (supervised) with labeled data specific to a task

WORD EMBEDDING TECHNIQUES



More specific technical details… at following workshops

EIFFEL Overview
LIBRA’s Role
Technical/NLP Introduction
LIBRA’S Technical Tasks
Challenges
Impact

TASK 1: NATURAL LANGUAGE PROCESSING (NLP)-BASED COGNITIVE SEARCH FOR GEOSS DATASETS

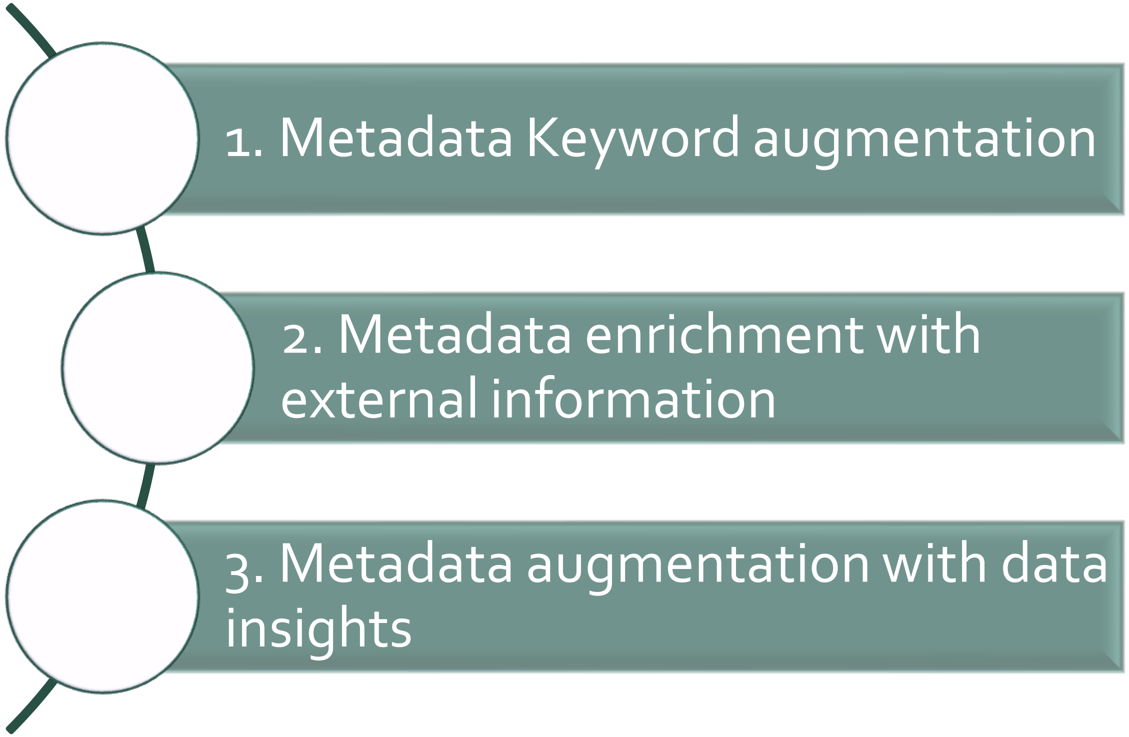
- Build the core of a search engine for GEOSS data based on AI-powered Natural Language Processing (NLP)
- The search engine will allow advanced query-related capabilities that add **semantic relevance** to search results & allows searching based on free text
- Domain specific
- Contextually rank the most relevant search results

Google for GEOSS data

TASK 2: METADATA ENRICHMENT MECHANISMS

- **Goal**: Diverse toolset for automatic metadata curation and augmentation, specialised to the GEOSS context

- **New Metadata DB** will be developed – not physically linked with GEOSS



1. Metadata Keyword augmentation

2. Metadata enrichment with external information

3. Metadata augmentation with data insights

EIFFEL Overview

LIBRA’s Role

Technical/NLP Introduction

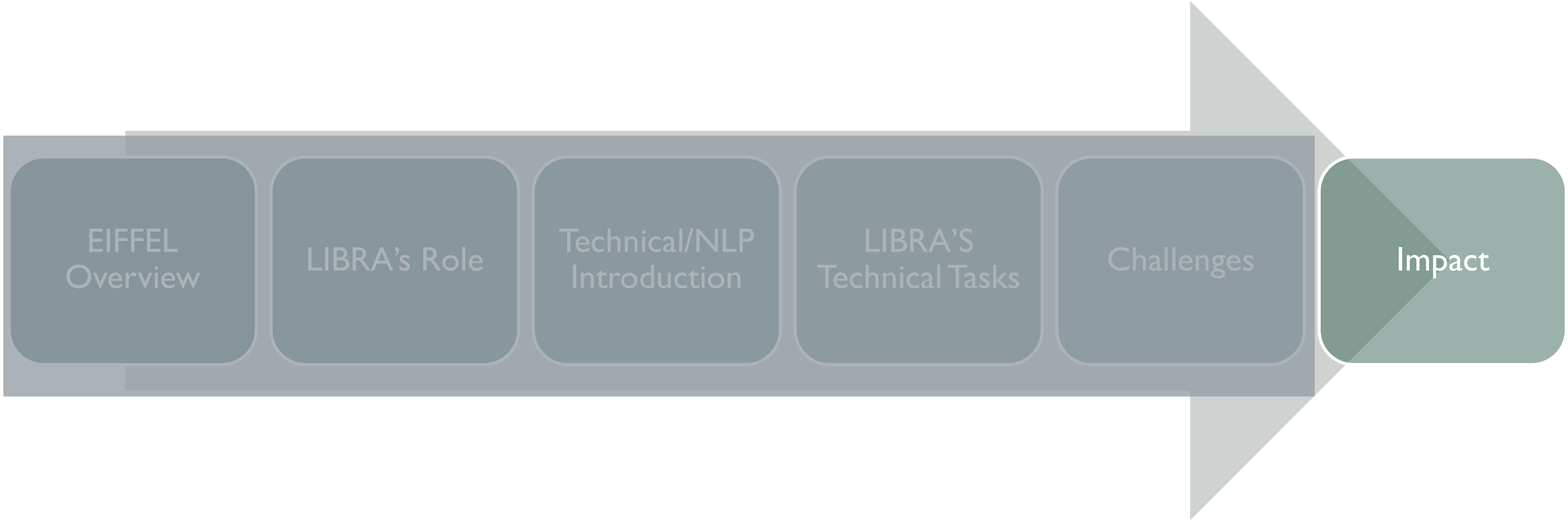
LIBRA’S Technical Tasks

Challenges

Impact

CHALLENGES FOR LIBRA

- **Very large and unstructured corpus** of geo-science documents.
- Current metadata not sufficient / lack a keyword list.
- GPT-3 produced the equivalent **of 552 metric tons of carbon dioxide** during its training. That's the same amount that would be produced by driving 120 passenger cars for a year.
- CC Applications depend on the quality of augmenting GEOSS data exploration activities.
- Leading role of WP: Coordination of many teams.


EIFFEL Overview
LIBRA's Role
Technical/NLP Introduction
LIBRA'S Technical Tasks
Challenges
Impact

IMPACT ON LIBRA

- Deep dive into various NLP tasks:
  - Context-aware Word and document embeddings.
  - Question-Answering.
  - Document Ranking.
  - Information Retrieval.

- Get accustomed with SotA models, such as Deep Bidirectional transformers, Generative pre-trained transformers.

- Make Libra's work known to the expert communities.

Thank you!

Questions?