

CLIMA-GAP

What Makes a Successful Climate Service?: Methodology Description

Paula Checchia (paula.checchia@bsc.es) - Barcelona Supercomputing Center (BSC-CNS)



1 PROJECT CONTEXT & DEFINITIONS

CLIMA-GAP is a research initiative aimed at identifying the **characteristics that contribute to the success of climate services**.

- Climate services:** Climate services are the provision of climate information to support decision-making. *Imagine a climate service like a pizza.*
- X criteria:** Characteristics that can be found in climate services. *The criteria represent the ingredients used in the pizza.*
- Y indicators** Success indicators. *The success indicators tell us whether the recipe turns out tasty.*

18 criteria defined by the Climateurope2 project

10 indicators defined through literature review

METHODOLOGY

- Selection of climate services case studies
- Collection of documents for each case study
- Development of a RAG system to evaluate whether each criterion and indicator is met in the case studies, and to explain why

2 COLLECTION OF DOCUMENTS FOR THE CASE STUDIES

We focused on EU-funded projects mentioning climate services available through CORDIS (the EU's Research & Innovation results platform).

CORDIS
Provides an API with links to project materials. We used project deliverables.

Search "climate service*"
112 results

API request
Excel with links to documents

Filtered by:
End date < Marh 2025 & >=10 deliverables
54 projects

Python script
Automatic download of PDFs (named by deliverable ID)

Why project deliverables?

- Publicly accessible
- Standardized structure
- Allow comparative analysis at scale

3 RAG SYSTEM TO ANALYZE CRITERIA & INDICATORS

RAG SYSTEM ARCHITECTURE

1 Database creation

PDFs → text → chunks → vectors

Stored in ChromaDB

2 Retrieval

Hybrid search (semantic + keywords)

Top 10 chunks re-ranked

3 Generation

LLM (Qwen3-32B) uses query + retrieved chunks

Returns YES / NOT SPECIFIED + quotes

EXAMPLE PROMPT

Purpose: Extract textual evidence from the provided documents to assess whether a given criterion is met.

Criterion: [Insert criterion name and definition]

Instructions:

- Indicate if the criterion is met by answering "YES" or "NOT SPECIFIED".
- Support your answer with exact text excerpts from the documents.
- Include a concise summary.

EXAMPLE OUTPUT

```
{  
  "result": "YES",  
  "justification": [  
    "... quote 1 ... - [doc1] - [p.12]",  
    "... quote 2 ... - [doc2] - [p.7]"  
  ],  
  "summary": "The document describes..."  
}
```

Unstructured PDF extraction

LangChain Text chunking

Yake/KeywordExtraction Keyword extraction

sentence-transformers / paraphrase-multilingual-mpnet-base-v2 Embeddings

ChromaDB Vector database

cross-encoder/ms-marco-MiniLM-L6-v2 Re-ranking

Qwen3-32B LLM for inference

4 VALIDATION

Selection: 5 testing projects & 2 experts per project

Goal: Measure both the quality of the evidence retrieved and the reliability of final answer.

For each criterion/indicator, the experts were tasked with the following:

- Provide their project knowledge:** indicate whether, based on their own experience, the criterion/indicator was met in the project.
- Evaluate the chunks:** indicate whether each chunk retrieved contained relevant evidence to address the criterion/indicator.
- Provide their own answer based on the chunks:** using only the provided chunks, indicate whether the criterion/indicator was met or not specified.
- Read the RAG's final answer**
- Provide additional comments**

5 METHODOLOGY CONCLUSIONS

Retrieval

Hit	76,79%
NDCG	61,25%

Many retrieved chunks are non-relevant; however, at least one (often two) relevant chunks are typically retrieved and tend to rank near the top.

LLM

Recall	87,10%
F1_score	82,99%

The LLM is good at recognizing whether information is relevant or not, and it responds appropriately based on the chunks it receives

Adjustments to the retrieval process may be beneficial for future studies

Deliverables may not be the best source of information for this type of question