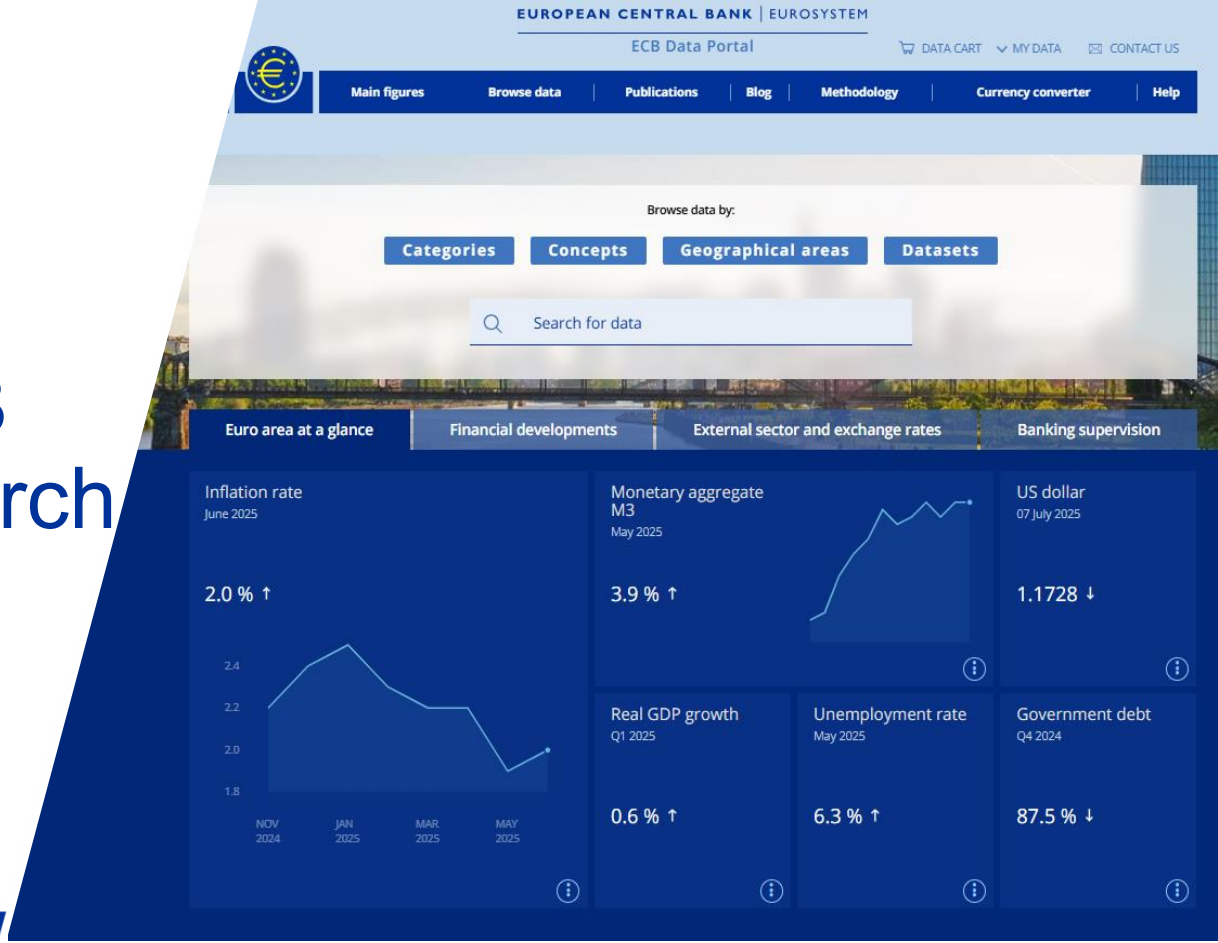
EUROPEAN CENTRAL BANK

EUROSYSTEM

SDMX and ECB Data Portal Search Engine



SDMX Global Conference, 2025
Rome, Italy

Ole Sorensen
Gabor Horvath
Zlatina Hofmeister

ECB Data Portal Search Engine

- ECB Data Portal (EDP) search engine
  - Uses OpenSearch
  - Input: SDMX metadata – on time series and sibling's level

- Following user feedback & usage statistics analysis, three main pain points related to EDP search were identified

  - **Information overload:** the portal disseminates ~4 million time series, Search returns too many time series, difficulties with narrowing down results.
  - **Metadata quality:** time series' titles are not descriptive enough and are using too technical language
  - **Search accuracy:** search engine improvements are essential, such as better ranking and accuracy

Launch of the **experimental projects** to address these issues

How do search engines work?

Three main functionalities

How do EDP search engine work?

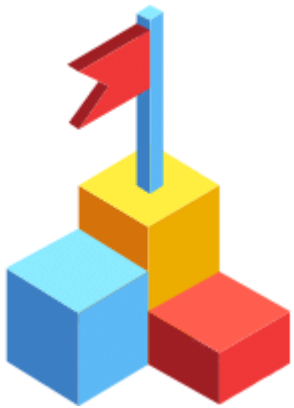
metadata

Series key	ICP.M.U2.N.011510.4.ANR
Last updated	17 July 2025 11:00 CEST
Unit	Percentage change
Frequency	Monthly (M)
Reference area	Euro area (changing composition) (U2)
Collection indicator	Average of observations through period (A)
Decimals	One (1)
Source	EUROSTAT
Adjustment	Neither seasonally nor working day adjusted (N)
Title	HICP - Butter
Title complement	Euro area (changing composition) - HICP - Butter, Annual rate of change, Eurostat, Neither seasonally nor working day adjusted
Unit multiplier	Units (0)
Tags	butter, food, HICP
Usage statistics	100
Publications	2

searching for

butter price

index



Experiment 1 objectives

Objective: Boost relevance, precision, and effectiveness of EDP search engine

Experiment 1

- Refines query building with improved selective filters and positional boosting.
- Involves reassessing existing filters, optimizing boosting, and introducing tailored query templates.
- Delivers more relevant search results without complex AI/ML changes.

What does that mean?

- Boost the important metadata
- Chose optimal boosting weight

Series key	ICP.M.U2.N.011510.4.ANR
Last updated	17 July 2025 11:00 CEST
Unit	Percentage change
Frequency	Monthly (M)
Reference area	Euro area (changing composition) (U2)
Collection indicator	Average of observations through period (A)
Decimals	One (1)
Source	EUROSTAT
Adjustment	Neither seasonally nor working day adjusted (N)
Title	HICP - Butter
Title complement	Euro area (changing composition) - HICP - Butter, Annual rate of change, Eurostat, Neither seasonally nor working day adjusted
Unit multiplier	Units (0)
Tags	butter, food, HICP
Usage statistics	100
Publications	2

Experiment 2 objectives

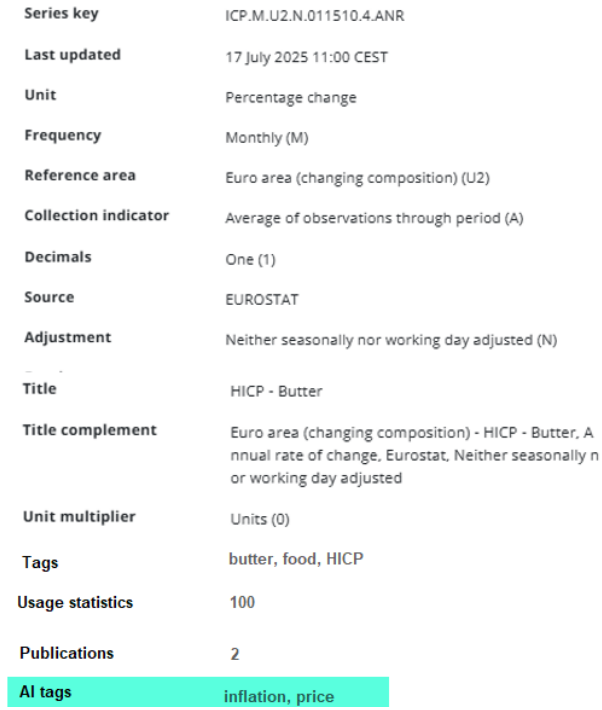
Objective: Boost relevance, precision, and effectiveness of EDP search engine

Experiment 2

- Applies AI-driven metadata extraction (tags, keywords, seasonality) to enrich metadata
- Uses clustering algorithms (K-Means, Hierarchical) for pattern detection and granular filtering.
- Automates boosting and intelligent filtering for refined search precision.

What does that mean?

- Add AI generated metadata in addition to existing metadata



Series key	ICP.M.U2.N.011510.4.ANR
Last updated	17 July 2025 11:00 CEST
Unit	Percentage change
Frequency	Monthly (M)
Reference area	Euro area (changing composition) (U2)
Collection indicator	Average of observations through period (A)
Decimals	One (1)
Source	EUROSTAT
Adjustment	Neither seasonally nor working day adjusted (N)
Title	HICP - Butter
Title complement	Euro area (changing composition) - HICP - Butter, Annual rate of change, Eurostat, Neither seasonally nor working day adjusted
Unit multiplier	Units (0)
Tags	butter, food, HICP
Usage statistics	100
Publications	2
AI tags	inflation, price

Experiment 3 objectives

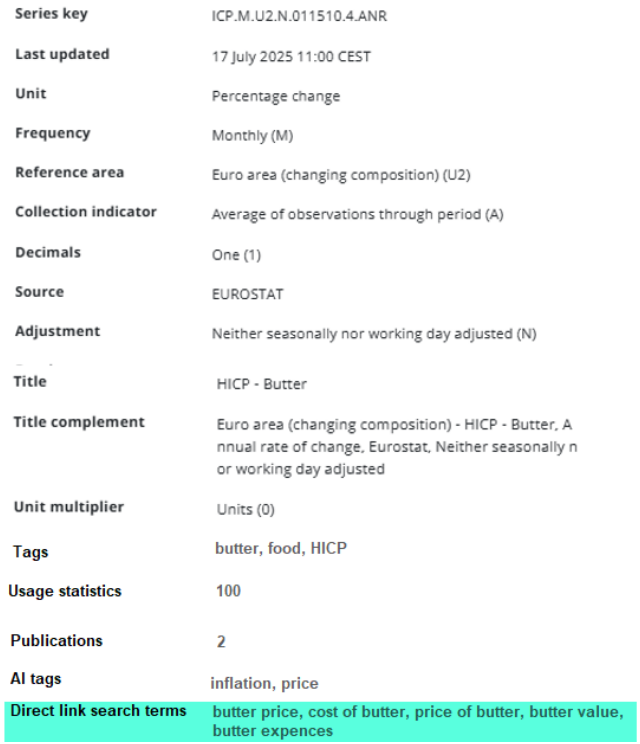
Objective: Boost relevance, precision, and effectiveness of EDP search engine

Experiment 3

- Advances query expansion via optimized templates, stop word lists, and synonym files.
- Leverages AI for keyword extraction and aligns templates with dataset clusters.
- Implements a phased approach for testing, optimization, and validation.

What does this mean?

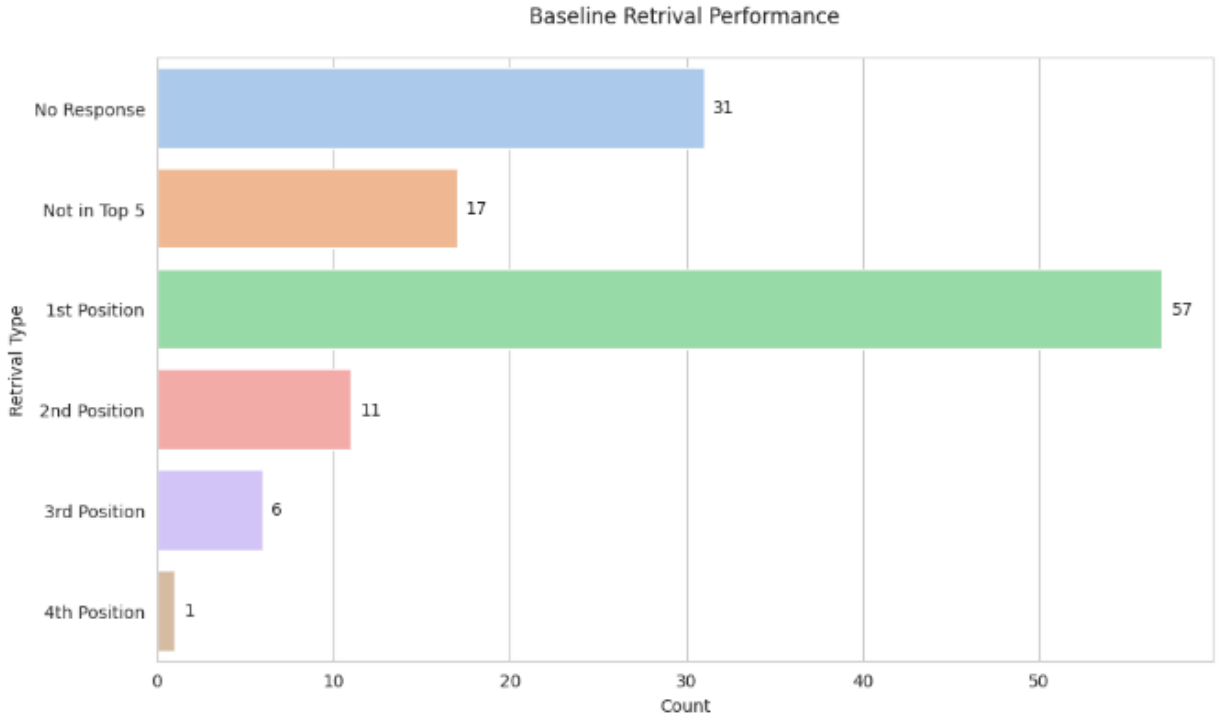
- Directly liking specific search terms to specific series/set of series/dataset



Series key	ICP.M.U2.N.011510.4.ANR
Last updated	17 July 2025 11:00 CEST
Unit	Percentage change
Frequency	Monthly (M)
Reference area	Euro area (changing composition) (U2)
Collection indicator	Average of observations through period (A)
Decimals	One (1)
Source	EUROSTAT
Adjustment	Neither seasonally nor working day adjusted (N)
Title	HICP - Butter
Title complement	Euro area (changing composition) - HICP - Butter, Annual rate of change, Eurostat, Neither seasonally nor working day adjusted
Unit multiplier	Units (0)
Tags	butter, food, HICP
Usage statistics	100
Publications	2
AI tags	inflation, price
Direct link search terms	butter price, cost of butter, price of butter, butter value, butter expences

Accuracy* - current model – 60%

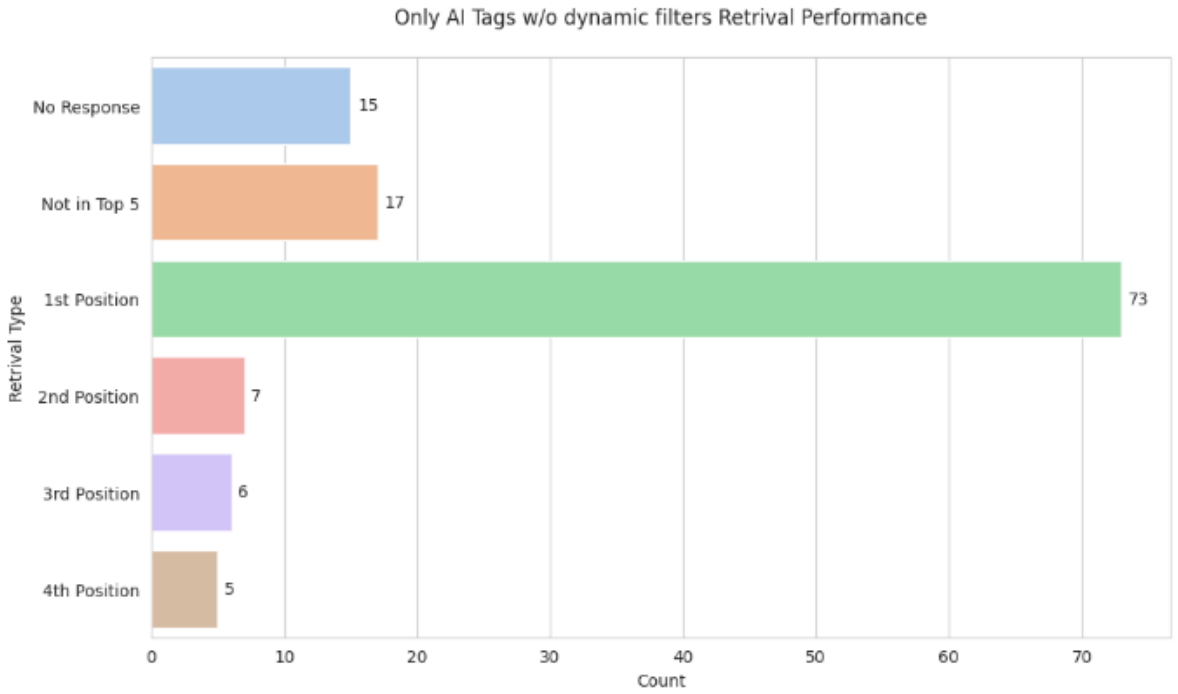
Based on sample of 123 search terms



*Accuracy is defined by having the expected result in top 4 positions of the search result

Accuracy* – Experiment 2 – 73%

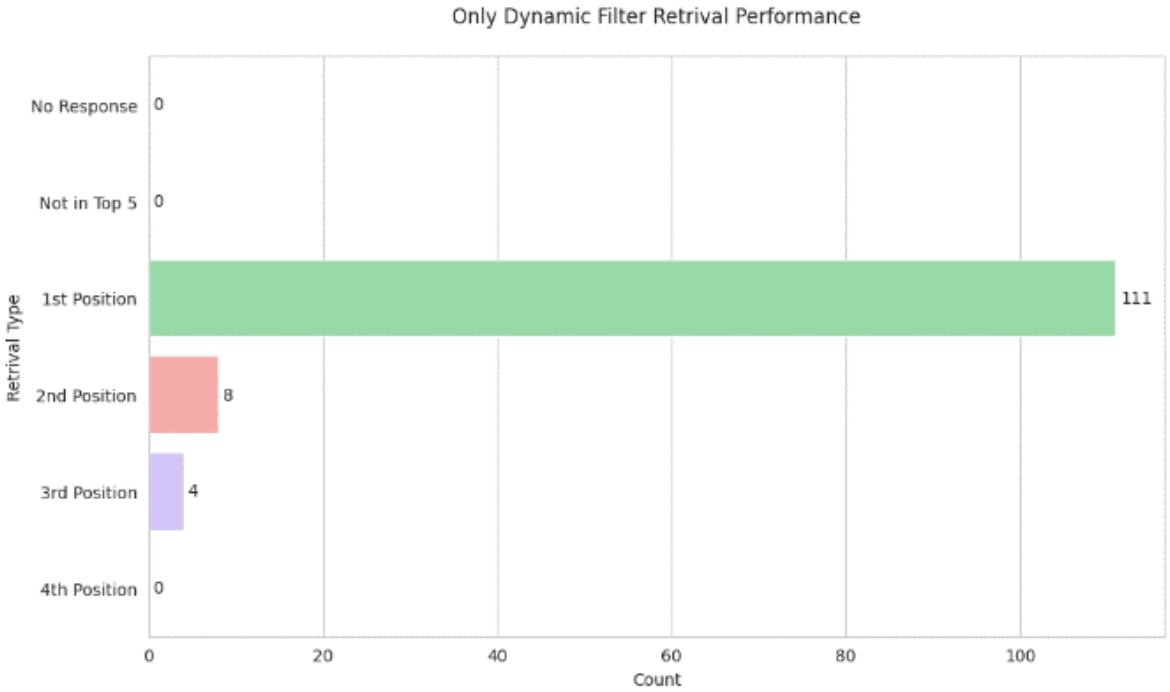
Based on sample of 123 search terms



*Accuracy is defined by having the expected result in top 4 positions of the search result

Accuracy* – Experiment 2 and 3 combined – 100%

Based on sample of 123 search terms



*Accuracy is defined by having the expected result in top 4 positions of the search result

Conclusion and Next Steps

🔑 Key Takeaway

- **Increased Accuracy**: A complex task which requires manual expert input from which the model learns
- **Small sample:** accuracy increase should not be generalized as the sample is very small
- **More work to be done as part of EDP 2.0**: more sophisticated query expansion methods and learning to rank models will be considered