

Commonwealth of Tessaly 2026 — GBP 2043 Notes

CDSBench #120 · What is the maturity date of the instrument? Answer as YYYY-MM-DD.

Commonwealth of Tessaly · Sovereign · Restructuring 2026-01-10 · Fact extraction · Extraction · Ground truth: deterministically verified

VERIFIED OUTCOME

2043-01-10

Stated in the maturity condition.

Supporting sections: DOC-1#S3

WHAT HAPPENED

All 4 participants reached the verified outcome.

Decided by analyst: The maturityDate is 2043-01-10.

MODEL COMPARISON

PARTICIPANT	ANSWER	RESULT	CONFIDENCE	EVIDENCE SCORE	PRIMARY FAILURE
Claude	2043-01-10	Correct	75%	100%	—
GPT	2043-01-10	Correct	92%	0%	Evidence selection failure
Gemini	2043-01-10	Correct	85%	0%	Hallucinated evidence
Mastermind (reference)	2043-01-10	Correct	75%	100%	—

GPT

GPT: Evidence selection failure

Correct answer but cited evidence does not match the supporting sections (evidence score 0.00).

Model summary: The maturityDate is 2043-01-10.

Cited: DOC-1#S1

GEMINI

Gemini: Hallucinated evidence

1 cited section id(s) do not exist in the case record.

Model summary: The maturityDate is 2043-01-10.

Cited: DOC-1#S1, DOC-9#S9 (1 non-existent)

EVIDENCE

Decisive sections of the record

- DOC-1#S3 — GBP 2043 Notes — Terms and Conditions (excerpt), Maturity: The GBP 2043 Notes mature on 2043-01-10. Interest is payable semi-annually in arrear.

METHODOLOGY

A response is correct when its normalised answer equals the verified label. Evidence score is the F1 between cited section IDs and the supporting sections; cited IDs that do not exist count as hallucinated. Abstentions are correct only when the label calls for them. Ground truth is verified by two independent extractors plus an evidence check; no model grades itself.

Full Case Replay with documents and every model's reasoning: <https://cdsbench.com/cases/commonwealth-of-tessaly-2026/gbp-2043/maturity-date>