

Complexity Calculation

Complexity Score Formula

The complexity score is calculated using this SQL formula

```
ROUND(  
    (a.word_count * 0.7) +  
    (COUNT(c.id) * 1000 * 0.3)  
) as complexity_score
```

Breaking Down the Formula

1. Word Count Component (70% weight)

- Takes the agency's `word_count` field
- Multiplies it by `0.7` (70% weighting)
- This represents the **textual complexity** - more words = more regulatory content

2. CFR References Component (30% weight)

- Counts the number of CFR references (`COUNT(c.id)`) for each agency
- Multiplies by `1000` to scale it up (since CFR reference counts are typically small numbers)
- Then multiplies by `0.3` (30% weighting)
- This represents the **regulatory breadth** - more CFR references = more regulatory domains

3. Final Calculation

The formula combines both components:

Complexity Score = (Word Count × 0.7) + (CFR References × 1000 × 0.3)

Example Calculation

Let's say we have an agency with:

- **Word Count:** 50,000 words
- **CFR References:** 15 references

Calculation:

- Word component: $50,000 \times 0.7 = 35,000$
- CFR component: $15 \times 1000 \times 0.3 = 4,500$
- **Total Complexity Score:** $35,000 + 4,500 = 39,500$

Why This Approach?

70% Word Count Weight

- Word count is a direct measure of regulatory volume
- More words typically mean more detailed regulations, requirements, and complexity
- It's the primary indicator of how much regulatory content an agency manages

30% CFR References Weight

- CFR references indicate how many different regulatory domains an agency covers
- An agency with references across many CFR titles has broader regulatory reach
- The 1000× multiplier ensures CFR references have meaningful impact despite being smaller numbers

What It Measures

This score attempts to quantify:

- **Regulatory Volume** (how much regulatory text)
- **Regulatory Breadth** (how many regulatory areas)
- **Overall Regulatory Burden** an agency represents

Agencies with high complexity scores are those that either have extensive regulations (high word count) or regulate across many domains (many CFR references), or both.