

TED data transparency using AI

Harnessing AI approaches to solve practical problems

David Dowey

Publications Office of the European Union



TED AI Proof-of-concept

Design	AI experiments based on stakeholder needs
--------	---

Build	Demonstrators to show and validate results
-------	--

Explore	Development options, resources, implementation methodologies for operationalisation
---------	---

Conclude	Produce a study report with recommendations
----------	---

Definition of the needs for AI



Assisting

Co-pilot: anticipate and propose

Ensure completeness and coherence



Streamlining

Efficiency throughout the publication process

Matching of similar needs



Verifying

Trustworthiness of data and reliability of results

Detect unlawful behaviour and fraud

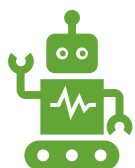
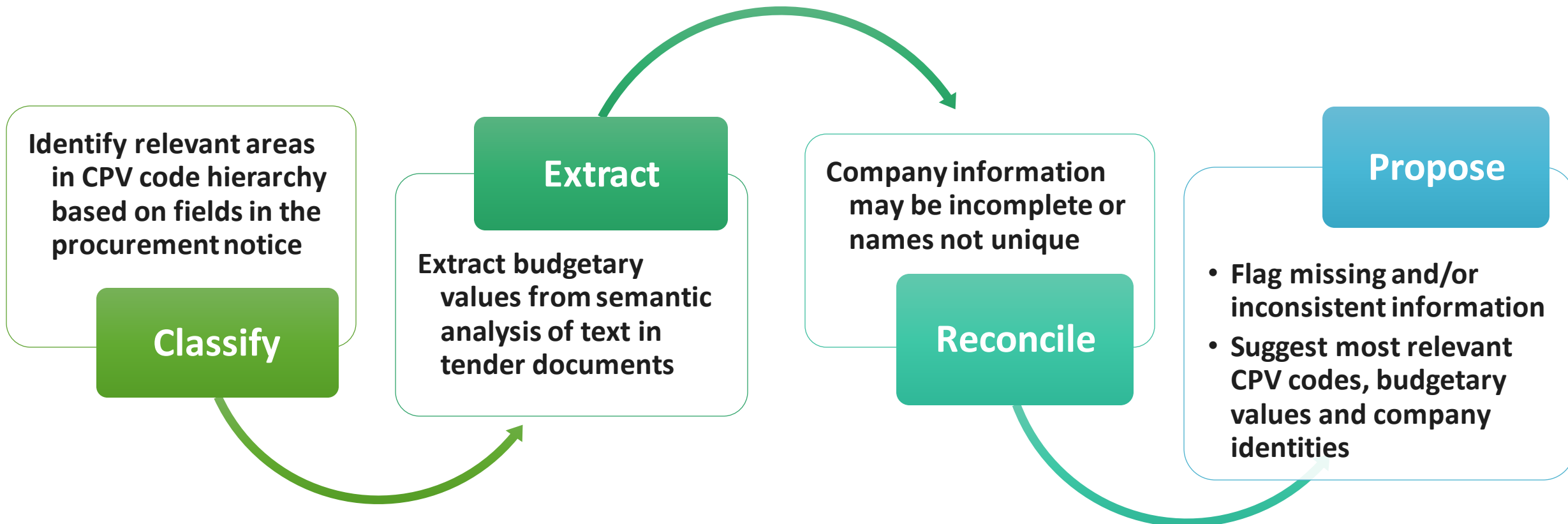


Understanding

Enable powerful queries

Federation and aggregation of data

Assisting users as a co-pilot



Virtual assistant: Answer questions and guide users step-by-step

Streamlining: useful AI techniques



Are CPV codes the most effective way to classify procurement?

- CPV classifier extended to full tender documentation
- Clustering techniques with topic modelling approaches
- Exploiting the hierarchy in the CPV taxonomy

How do similar tenders distribute in terms of value?

- Identify similar tenders based on context/language of the description text & by number of lots
- Can AI help to benchmark values for assessing tenders?

How can patterns of buyers be identified?

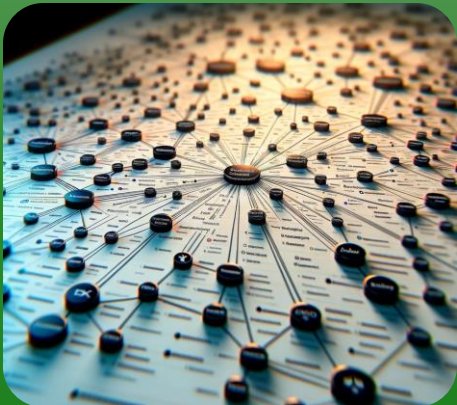
- Similarity of calls for tender; with a domain, over time
- To which contractor have most contracts been awarded, per buyer? Why is that?

Verifying to ensure trustworthiness



Ensure that TED publishes only notices from lawful buyers

- Matching of combinations of metadata with business rules
- Term search applied to “forbidden” words such as “Test”, “Sales”
- Gibberish text detection based on pretrained model
- Entity matching against a whitelist



Unique identification of companies, their ownership structure and interrelationships

- ORBIS dataset remapped to RDF and ingested into a knowledge graph
- Identify companies for further scrutiny – changes to organisational structure, business activities
- Connection between the procurement (contracts awarded) and companies

Understanding of aggregate, linked data



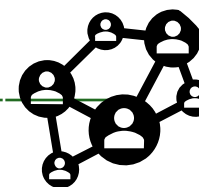
- AI-assisted pipelines to retrieve, process, homogenise and curate the data
- Aggregate TED data in order to derive statistics and to make data-driven decisions
- Query procurement data in relational DB
 - Connect information from contract notices and contract award notices
 - Top winners from a contracting authority
 - Amounts awarded over time
- Query company data in knowledge graph DB
 - Companies with inconsistent information, many changes, lack of independent oversight
 - Who obtained most contracts? Where?
 - Executive turnover in high-winning companies



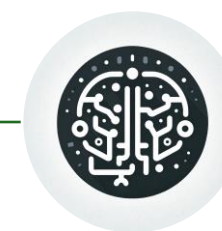
Future paths



Use longer text inputs & create gold standard validation data



Extend the company knowledge graph with more procurement information



Further investigate the use of LLMs and adapted prompting techniques



Implement reliable solutions in production

Thank you!

Any questions? Thoughts?



© European Union 2024

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Slides 6, 7 and 8 contain images generated by artificial intelligence with no third-party copyrights.

