

BI legal unit creation

Hybrid: machine learning and rules-based record matching

Isabela Breton
Dana Seman-Bobulska
Joanne Sheppard

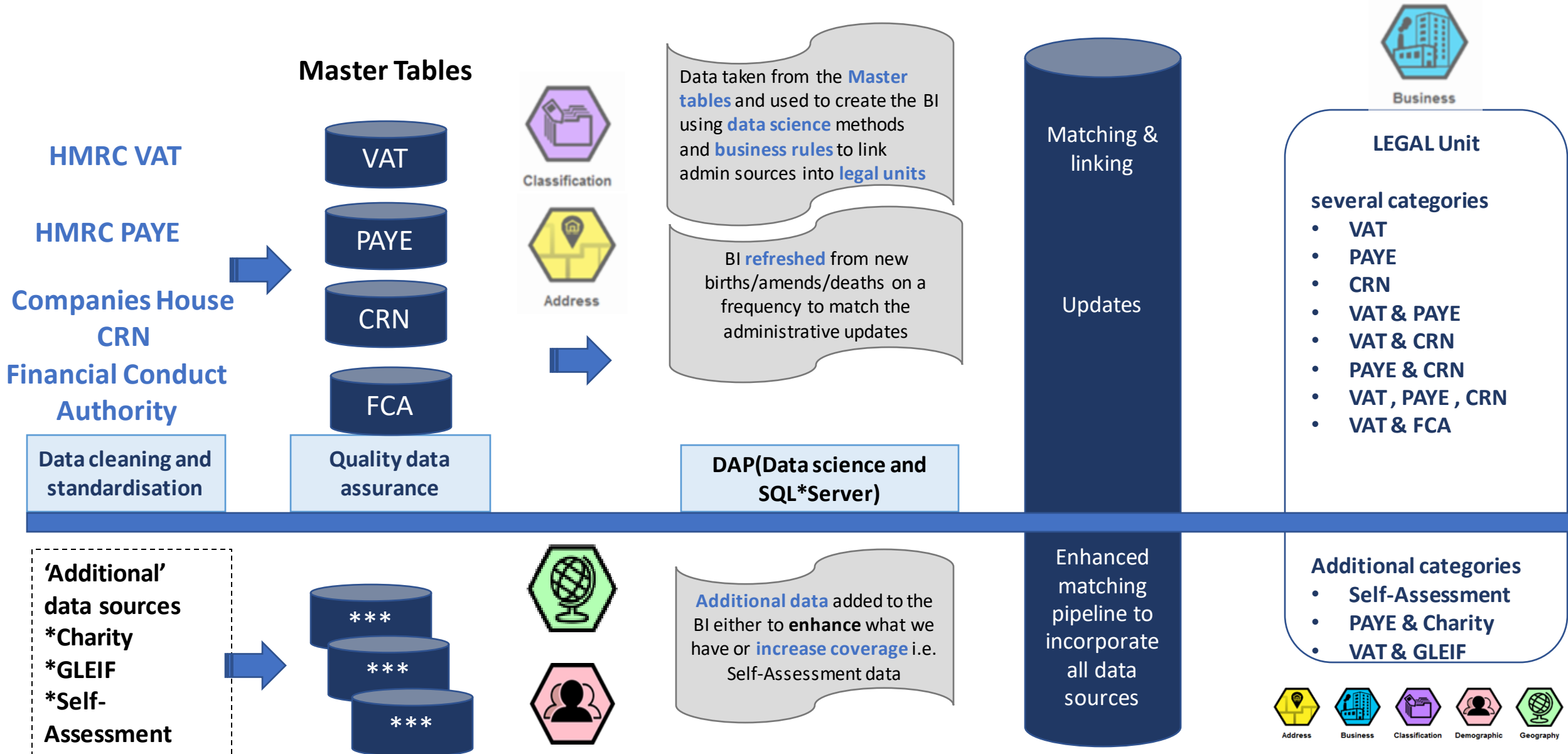
11/05/2023



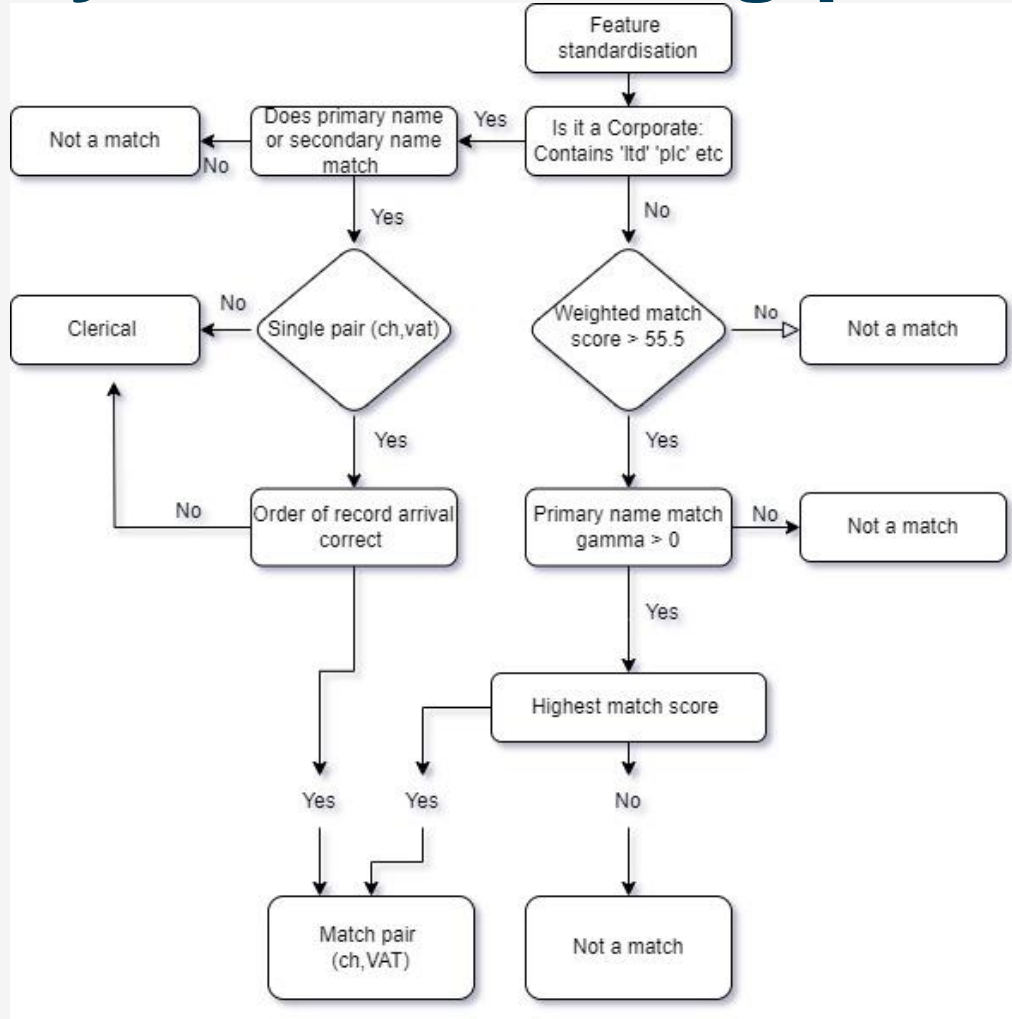
Overview

- Business Index
- Hybrid record matching approach
- Development approach
- Conclusions
- Future plans
- Questions

Business Index



Hybrid: matching process



Feature standardisation



Separate corporates



Rules based matching for corporates



Probabilistic matching for non corporates

Development approach



Rapid deployment



Logging



Feedback from customers

Conclusions



Hybrid: using Splink with a rules-based approach demonstrated has delivered high accuracy match rates across multiple datasets. This is the result of combining modern methods and extensive domain expertise.



Development approach: The accuracy improved steadily over time due to the early deployment of the hybrid pipeline and continuous monitoring and feedback from the domain expert

Future work



Front end: matching service development



More data: FCA, GLEIF, Charities



Methodology quality assurance review (ongoing)



Continuous improvement ensure tool remains cutting edge (via logging / clerical feedback)



Future platform: Google Cloud Platform migration

Questions