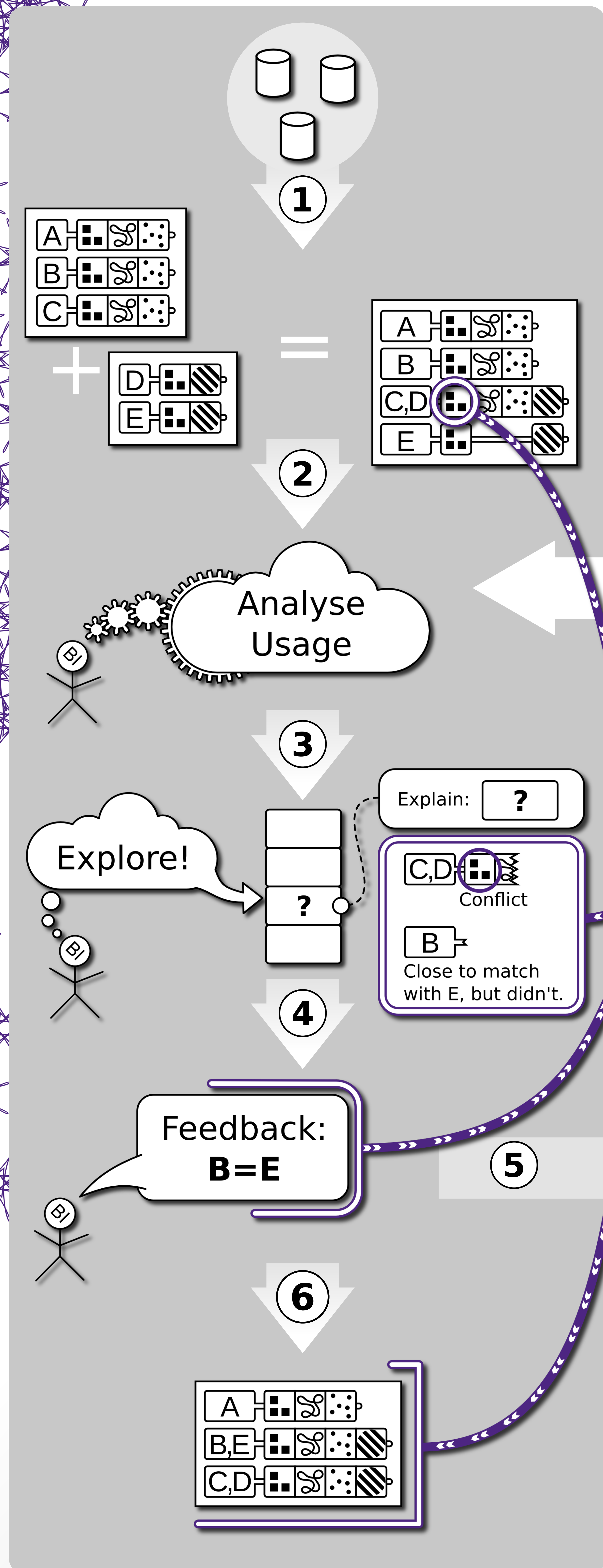
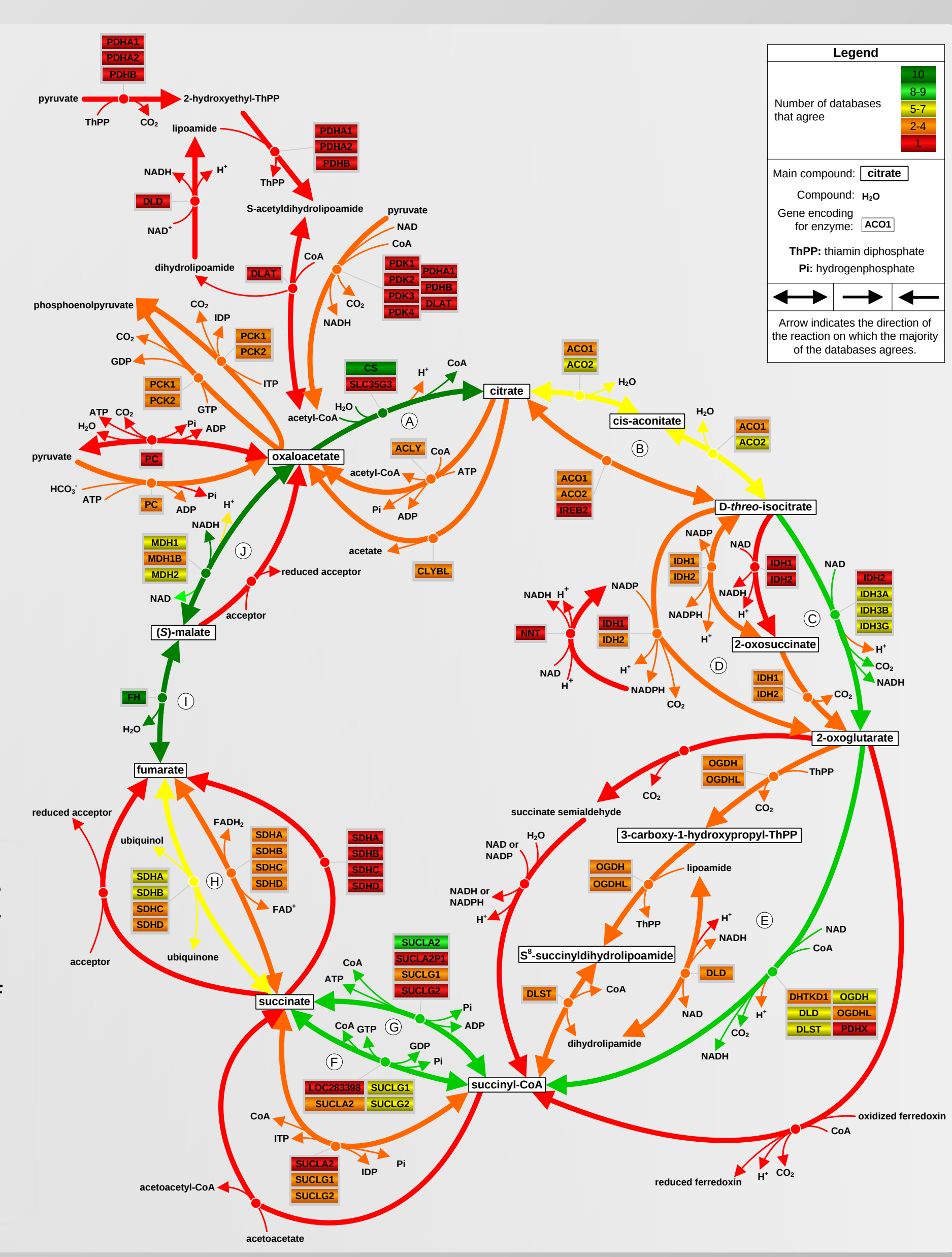


# PayDIBI | Pay-as-you-go data integration for bioinformatics

Brend Wanders • Paul van der Vet • Maurice van Keulen  
 <b.wanders@utwente.nl>



- 1 Combination of multiple data sources, each created with a specific purpose in mind. This first integration attempt will have inconsistencies and semantic mismatches.
- 2 The bioinformatician analyses the combination attempt to determine whether the current level of integration is good enough for their purposes.
- 3 Typical usage of integrated data by the bioinformatician is exploring and querying to gain insight into the structure and properties of the data. The system can explain the results of a query.
- 4 Based on their analysis and exploration, the bioinformatician can issue feedback and improved integration rules to the system. These rules are then stored in the personal knowledge base.
- 5 A new combination attempt is created in the next iteration, allowing the bioinformatician to continue refining their rules. This allows the bioinformatician to spend effort only when necessary.
- 6 When the integration level is to the bioinformatician's liking, they can stop iterating over the analyse-explore-feedback loop and use the integrated data. At a later point, the bioinformatician can come back to step 5 to continue refining the integration.



Union of the descriptions of the TCA cycle given by 10 pathway databases. Colors indicate the level of agreement on a reaction and on the genes.  
 Image courtesy of M.D. Stobbe et al. "Improving the description of metabolic networks: the TCA cycle as example". In: The FASEB Journal 26.9 (2012), pp.