

# Poster Application for BBC 2012

**Name:** Brend Wanders <b.wanders@utwente.nl>

**Affiliation:** Databases Group, Faculty of Electrical Engineering, Mathematics & Computer Science, University of Twente

## Abstract

**Background** Scientific research in bio-informatics is often data-driven and supported by biological databases. In a growing number of research projects, researchers like to ask questions that require the combination of information from more than one database. Most bio-informatics papers do not detail the integration of different databases. As roughly 30% of all tasks in workflows are data transformation tasks, database integration is an important issue.

Integrating multiple data sources can be difficult. As data sources are created, many design decisions are made by their creators.

**Methods** Our research is guided by two use cases: homologues, the representation and integration of groupings; metabolomics integration, with a focus on the TCA cycle.

**Results** We propose to approach the time consuming problem of integrating multiple biological databases through the principles of ‘pay-as-you-go’ and ‘good-is-good-enough’. By assisting the user in defining a knowledge base of data mapping rules, trust information and other evidence we allow the user to focus on the work, and put in as little effort as is necessary for the integration. Through user feedback on query results and trust assessments, the integration can be improved upon over time.

**Conclusions** We conclude that this direction of research is worthy of further exploration.