# BioFuice: A decentralized approach to integrate molecular biological annotation data

Molecular biological annotation data is continuously being collected, curated and made accessible in numerous public data sources. Many applications require annotation data from different sources. Integrating such highly diverse data is a major challenge in bioinformatics and often error-prone for scientists. Traditional database approaches based on a global schema are too rigid and not scalable due to the difficulty to find a suitable centralized schema and to map the diverse data to it.

We present the BioFuice approach for interconnecting and integrating annotation data from different autonomous sources. It is based on a decentralized peer-to-peer-like infrastructure. We directly utilize instance (object)-level correspondences between different sources which are often already available in the sources in the form of web links. Sets of such correspondences represent mappings between sources which describing objects of different types, such as gene, protein, and their function. Mappings are also assigned a semantic mapping type within a domain model. To process objects and mappings we have devised a set of high-level operators. They can be used within script programs to combine and analyze data from different sources. For instance, we can use a script to retrieve all proteins and pathways that are affected by a set of differently expressed genes found in microarray-based experiments. This data then can be utilized for further analysis in other applications. Currently, BioFuice integrates data from more than 10 sources for various research projects ranging from analysis of microarray data to the detection of non-coding RNAs.

**Contact:**

Toralf Kirsten, Erhard Rahm
Interdisciplinary Centre for Bioinformatics (IZBI)
Härtelstraße 16-18
04107 Leipzig