# Accessing chemical and biological datasets through SPARQL endpoints

Galgonek Jakub1

1 Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Flemingovo náměstí 542/2, 160 00 Praha 6, Czech Republic

Many of medium to large scale biological and chemical datasets are internally stored in relational databases. This approach makes it easy to develop a dedicated web server that presents a dataset and supports data querying. Unfortunately, such a way is not usually interoperable, and it can be difficult to combine the dataset with other ones or to query multiple data sources uniformly. To address this gap, some of these datasets use the Resource Description Framework (RDF) to export their data in interoperable formats. To increase interoperability even more, some datasets support querying their data by SPARQL, the query language for RDF data. If a dataset is originally stored in a relational database, there are two basic approaches. In the first one, data are exported to an RDF form and stored in a native RDF storage supporting SPARQL querying. The disadvantage of this approach is that the data are stored twice, or migration of the used database technology is needed. The second approach is to keep the data in the relational database and use a system that allows mapping the relational data to the RDF form. This mapping is used by the system to translate incoming SPARQL queries to equivalent SQL queries that are then evaluated by the relational databases. In our work, we examine different ways how these mappings can be designed, and we compare them with each other and with the native solution. We also compare several technologies, that are typically used for these purposes in the field of biology and chemistry.