# Digital Dictionary Database for Slovenian: unstructured, semi-structured and structured data in modern lexicography

Simon Krek (Jožef Stefan Institute, Slovenia)

Lexicography in 21st century is faced with many challenges, from the opportunities arising from the increasingly versed Artificial Intelligence, to the democratisation of knowledge through open collaboration and citizen science projects. However, it seems that the very basic *raison d'etre* of lexicography persists - to offer as much knowledge about the vocabulary of a particular language, and its interaction with other languages, as possible. The real challenge in the beginning of the 21st century is how to organise this knowledge which comes in different forms, as unstructured data (text, corpora), semi-structured data (XML, JSON etc.) or structured data (relational databases). Traditionally, from the beginnings of their use in 1980s, lexicography considered corpora as detached from lexicographic description, as a source consulted at a given point in time, not intrinsically connected with lexicographic resources, with such a connection ultimately enabling dynamic (semi-)automatic monitoring of language use in lexicographic fashion (identification of lexical units, sense division, explanation, exemplification etc.). Now, with the disappearance of book-driven organisation of lexicographic data, there is no need for compartmentalisation of data types. The new challenge is how to monitor, what to monitor, how to store lexicographically organised information, and how to visualise it. The organisation of Digital Dictionary Database (DDDS) for Slovenian tries to provide an answer for some of these challenges. In the talk, we will present DDDS data model, our tools for monitoring language use in real time, and visualisation of various types of data: monolingual data, bi- or multilingual data, corpus data, multi-modal data etc.