December 2024

LIKAT chemist receives award for digital data management and AI in catalysis

For the third time, the Catalysis Consortium of the National Research Data Infrastructure (NFDI) has presented its award for merit in the implementation of the FAIR principles in scientific data management. The "NFDI4Cat - Digital Chemist Award" 2024 went to Dr. David Linke from the Leibniz Institute for Catalysis Rostock, LIKAT, as announced at the annual meeting of the NFDI4Cat in November. The FAIR formula defines standards for the handling of research data according to these criteria Findability, Accessibility, Interoperability (data compatibility) and Re-Usability (reusability).

For the past four years, a team led by David Linke has been developing programs as part of the NFDI4Cat consortium that enable practically any laboratory to make its catalysis research data available digitally. Such data, which is accessible to everyone in the scientific community, is then also used to train AI models.

The data is the complete documentation of the procedure for all experiments, which can run into the hundreds for a single publication. Catalysis researchers usually only publish a fraction of the data in their papers, as David Linke says. "But there is ten to 50 times the amount of data behind it. For example, on experiments that were unsuccessful even though they were carried out correctly." And this is also important for feeding and training AI.

The tools that David Linke is developing as part of the National Research Data Infrastructure (NFDI) make it easier for chemists to prepare and document this data in such a way that machines can read it unambiguously and exchange it with each other. "This criterion is, so to speak, hidden behind the 'I' in the FAIR formula, interoperability, and it also represents the greatest challenge of our mission," says David Linke. Developing a precise vocabulary for this was his most important achievement, as even representatives of the so-called exact sciences often use the technical terms differently.

In the future, applications of artificial intelligence in science could benefit considerably if such models were to serve as "scientific experts" alongside the currently hyped and also completely differently constructed language AI solutions, the so-called *large language models.* This role can be assumed by so-called knowledge graphs, which can represent relationships precisely and integrate the precise vocabulary developed by David Linke and his colleagues, among other things.

What are the benefits of an AI-compatible pool of research data? According to Dr. Linke, it makes research much more effective. "I can create comprehensive cross-connections to my topic and see, for example, where other laboratories have left experimental gaps that could become interesting with my level of knowledge." The pool of knowledge that can be used worldwide is simply getting bigger.

The NFDI e.V. was founded in October 2020 by the federal and state governments with precisely this goal of digitization in the research data sector. The work is scheduled to run for a total of ten years, i.e. until 2030. The NFDI4Cat consortium was one of the association's first consortia to be founded in the field of chemical catalysis and, together with Chemistry Europe, initiated the Digital Chemist Award endowed with 1,000 euros.

Contact person:

Dr. David Linke

(Head of department "Catalyst Development and Reaction Technology")

David.Linke@catalysis.de

0381 1281 414