



Beitrag ID: 11

Typ: Talk

## **Self-assessment for FAIR data publication: empowering researchers to improve dataset quality before submission**

The CoreTrustSeal certified institutional data repository RDR, built on Dataverse, is central to KU Leuven's efforts to support FAIR data publication. Since its launch in 2022, the growing number of dataset submissions has highlighted the need for an efficient, transparent, and consistent curation workflow. To address this, the RDR team developed an open-source review dashboard that integrates with Dataverse and streamlines the curation process.

Initially designed to optimize the review workflow, the dashboard's second iteration introduced Python-based automated checks for systematic quality assessment. These checks validate metadata completeness and consistency while flagging issues such as missing PIDs, unclear licensing, insufficient metadata, or absent README files. Crucially, automation complements rather than replaces human judgement: curators can override or contextualize outcomes, ensuring nuanced interpretation remains part of the process.

Recurring metadata issues often surface only during curation, causing delays and additional review rounds. Building on the insights from the automated checks in the review dashboard, the RDR team is developing a self-assessment tool for researchers. This tool enables more complex pre-submission validation of draft datasets than is possible in the Dataverse UI and embeds FAIR-oriented guidance, including PID requirements, licensing clarity, consistent metadata, and documentation completeness. By providing concrete, and actionable feedback, it helps prevent common issues before formal review and supports the creation of more complete datasets.

The presentation will introduce the design principles and implementation of this self-assessment tool, highlighting the metadata checks and how feedback is presented to users. We will discuss how automated assessment assists researchers in fulfilling essential requirements while encouraging more complete metadata. Furthermore, we will reflect on key insights and challenges, offering guidance for institutions aiming to strengthen research support and enhance metadata quality for FAIR-aligned data publication.

### **ONLY WORKSHOPS - Proposed interaction format**

### **Alternative Track**

### **ONLY WORKSHOPS - Tentative audience**

### **ONLY WORKSHOPS - Maximum number of participants**

### **ONLY WORKSHOPS - Special technical requirements**

**Autor:** MARYNISSEN, Marleen (KU Leuven)

**Co-Autor:** Frau BLOEMEN, Dieuwertje (KU Leuven)

**Vortragende(r):** MARYNISSEN, Marleen (KU Leuven)

**Sitzung Einordnung:** TALK SESSION

**Track Klassifizierung:** HMC Conference 2026 Track Topics: 1. Metadata in Action: Embedding Quality and Context into Research Infrastructures