Contribution ID: **11**   Type: **not specified**

Workshop 7: Identifying Identifiers: Harnessing StrainInfo's tools to find and resolve strain designations in literature and metadata

Tuesday 30 September 2025 15:45 (2 hours)

Extracting meaningful biological entities (e.g. sequence accession numbers and species names) from unstructured text is an essential yet complex task. In microbiology, the identification is further complicated by the use of different designations (e.g. DSM 20543 and LMG 28910) for the same strain. Learn how to identify strain designations in literature and sequence metadata by leveraging StrainInfo's API and libraries. A workshop ideal for bioinformaticians, data scientists and researchers with a beginner-level understanding of Python, where participants will be guided through hands-on exercises. By the end of the workshop, attendees will have written a script capable of extracting strain identifiers from both literature and sequence metadata. This script will utilize the StrainInfo API to identify these strains and collect key information such as alternative designations, type strain status, taxonomy and more.

Presenter: LISSIN, Artur (DSMZ)

Session Classification: From Pipelines to Portals: Next-Generation Tools for Microbiome Analysis