

Mardi 24 mars 2026

11H00-12H00 – SALLE CHASSERAL, RTE DE LA CORNICHE 21 – 1010 LAUSANNE

« Deidentification of free-text documents for the responsible use of Artificial Intelligence: a use case in population-based cancer registration »

SPEAKER

Anisia Katinskaia joined the Swiss Data Science Center (SDSC) in August 2025 as a Senior Data Scientist in the Research team in Lausanne.

Anisia is an NLP researcher with a PhD in Computer Science from the University of Helsinki, Finland. She also holds a Master's degree in Computational Linguistics from the Russian State University for the Humanities. Anisia's research has primarily focused on applying NLP to education. During her time at the University of Helsinki, she worked on grammatical error detection and correction, text difficulty estimation, text simplification, low-resource NLP, model interpretability via probing, tokenization for language models, and the development of educational LLM-based agents. In SDSC, Anisia is working on biomedical NLP tasks, in particular, building knowledge graphs from unstructured scientific data and automatic cancer data processing.

ABSTRACT

Deidentification of free-text documents is a critical step in enabling the responsible use of Artificial Intelligence (AI) in sensitive domains such as population-based cancer registration. Cancer registries routinely collect large volumes of clinical narratives (e.g., pathology reports, hospital discharge summaries, imaging reports) that contain personally identifiable information (PII). To leverage these data for AI-driven research and public health insights, robust deidentification processes are essential to protect patient privacy while preserving data utility.

We present a deidentification approach for free-text pathology reports as part of the global AICADES project - a collaboration with Unisanté and the Swiss Data Science Center (SDSC) / École Polytechnique Fédérale de Lausanne (EPFL) - aimed at accelerating cancer-relevant data extraction for population-level surveillance. Our work addresses key challenges related to balancing privacy and data utility, rare disease re-identification risk, language variability, and governance.

Recommandé par la Société Suisse des Spécialistes en prévention et santé publique (SPHD) pour la reconnaissance de la formation continue.
Participer à la réunion WebEx.

[Lien WebEx](#)