EOSC | Federation

Scientific use case:
Multi-centric validation of AI models for prostate-cancer screening

EOSC Node | BBMRI-ERIC Health & Food | EOSC Node | Italy | EOSC Node | SURF The Netherlands

EOSC Node | Slovakia | EOSC Node | Poland | EOSC Node | Finland

Based on an interview with Andreas Türk, a data strategist at BBMRI-ERIC responsible for developing IT strategies within the biobanking research infrastructure community. His work includes overseeing digital tools, guiding development priorities, and coordinating data strategies. He joined BBMRI-ERIC in late 2022 and contributes to shaping its roadmap for handling sensitive biomedical data.

Advances in digital pathology are transforming cancer diagnosis by enabling high-resolution imaging of tissue samples and the application of artificial intelligence (AI) for clinical decision support. Prostate cancer, one of the most prevalent malignancies among men worldwide, represents a critical case for early and accurate diagnosis, as survival rates are strongly tied to timely detection and treatment. The objective of the **multi-centric validation of AI models for prostate-cancer screening (MCVAL)** use case, coordinated by the **BBMRI-ERIC EOSC Node**, is to create a secure environment in which AI models for prostate cancer screening can be validated using data from different hospitals. Rather than building new diagnostic systems from scratch, the project focuses on testing an existing model trained on whole-slide images and assessing how well it performs when applied to data processed elsewhere.

Problem addressed

AI models in digital pathology are often trained on data from a single hospital, limiting their generalizability. Differences in imaging protocols between institutions mean that a model validated in one place may not perform reliably in another. Moreover, sharing sensitive health data across hospitals for validation is difficult due to regulatory, ethical, and legal restrictions.

Technical solution

The use case leverages the EOSC Federation to establish a model for securing **validation of AI models using sensitive health data within compliant frameworks**. Ethical clearance is obtained

from partner institutions, while contractual agreements govern data transfers. The MCVAL use case begins with an AI model trained on data from Masaryk University's Masaryk Memorial Cancer Institute (CZ), and uses validation data from Medical University Graz (AT). High-performance computing resources are provided by the **Italian EOSC Node** at ICSC and, in the future, other EOSC Nodes. The long-term vision is to create catalogues of sensitive health data and computing infrastructures accessible to AI researchers across Europe, enabling them to validate and train models across diverse datasets and technologies, rather than repeating validation for each hospital individually.

Scientific outcomes

The scientific and societal impacts are twofold: first, **improving AI-assisted diagnostic tools** and potentially identifying **new diagnostic markers** for prostate cancer; second, contributing to earlier and more accurate detection of prostate cancer, thereby **improving treatment outcomes**. The timeline for tangible results is short-term, with initial demonstrations planned already in 2025, and with broader scientific and clinical impact expected in the following years. The use case contributes to better treatment outcomes and aligns with **Europe's Beating Cancer Plan**.

The use case also fosters collaboration between pathologists, AI researchers, hospitals, and biobanks, with patients ultimately benefiting from **improved diagnostics**. The project contributes to Open Science by fostering new collaborations between health providers, data scientists and AI engineers, enabling **model validation in secure environments**, and aligning with FAIR data principles. While AI models cannot always be openly shared due to ethical constraints, the infrastructure seeks to maximize accessibility within compliant frameworks by ensuring that the catalogues for sensitive health data and compute resources are FAIR..

Added value of the EOSC Federation

The EOSC Federation is central to the project, as it enables collaboration across a wider group of experts and provides access to computing resources and additional data sources that would not otherwise be available. EOSC Federation working groups, including those focused on sensitive health data, have connected the project with the **Slovak, Polish, Finnish and Dutch (SURF) EOSC Nodes**, expanding its data and compute resource base. The ultimate ambition is to transition from a single-use case to a sustainable service supporting a broad range of researchers and healthcare providers across Europe. By supporting the uptake of AI-based diagnostics, strengthening personalized medicine, and aligning with EU health policy goals, the use case can make a substantial contribution to innovation in Europe.