

Scientific use case: Federated analysis of pathogen genomes

eOSC Node | Slovakia

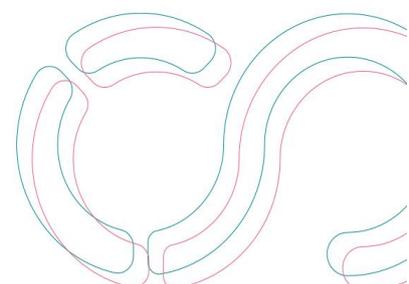
eOSC Node | Poland

This briefing is based on an interview with Jaroslav Budiš, a scientific officer at CVTI SR who is working on Slovakia's national EOSC Node. Budiš helped build the country's pathogen surveillance system that during the COVID-19 pandemic linked Slovak public-health laboratories, regional branches and university labs to harmonize analyses of national data, experience that now anchors his work on EOSC.

The **federated analysis of pathogen genomes** science case outlines a federated, cross-border capability for timely analysis of pathogen genomes that brings computation to the data instead of copying sensitive datasets across institutions. The objective is to shorten time-to-insight for outbreak detection, source attribution, and antimicrobial-resistance (AMR) surveillance while preserving data sovereignty and meeting European legal and ethical requirements. Experience from COVID-19 showed that sequencing at scale can transform public-health decision-making. Operationally, the effort starts with two neighbouring nodes of the EOSC Federation—the **Slovakian national node** providing workflows, datasets, computational infrastructure and domain expertise, and the **Polish national node** (via Poland's National Science Centre (NCN) and a scientific repository service) supplying key technical support and their own datasets.

Problem addressed

The use case comes in response to a central problem: many labs generate genomic data but keep them in **isolated databases**; when the data eventually reach public repositories, they're often delayed, inconsistently processed, and thus hard to harmonise across centres. A further gap is that **parallel application domains**—human health, veterinary and food safety, for example—study the same microbes in different contexts **but rarely connect their datasets**. The EOSC Federation offers a technical solution to connect these sources and analyse them with harmonized workflows. In practical terms, a clinician sequencing a patient's *Listeria*, for example, could rapidly compare it with *Listeria* from supermarket cheese to **trace the source of illness**.



Technical solutions

The geographical proximity of the **Polish and Slovakian Nodes** make them an ideal pair for a cross-border pilot. Technically, the system will leverage the **EOSC Federation's e-infrastructure backbone**. Identity and access control will rely on the EOSC Federation Authentication and Authorization Infrastructure (AAI), integrating with research identity providers to support **fine-grained authorization**, auditability and policy enforcement across Nodes. The immediate focus of the use case is to use already-public repositories to prove the technical foundation since timely clinical data are sensitive, embargoed and legally complex to share. Once trust is established, the use case envisions carefully governed links to authorities' private repositories. In other words, **"build the plumbing first"**, then invite data owners to experience the advantages of sharing data.

Scientific outcomes

Establishing a federation of trusted sites for sharing of sensitive data can **increase critical data sharing on the advance of AMR and improve cross-border surveillance of pathogen genomes**. The use case proposes unified, reusable workflows running across existing European infrastructures to **expand the view** beyond application domains and national borders, **improving outbreak tracing** and providing insights into **how pathogens spread and evolve**. The approach began with Salmonella, but in principle can support any organism sequenced with standard technologies such as Klebsiella and Listeria. The use case utilizes published workflows used in Slovak national surveillance—including **pipelines for respiratory viruses and antimicrobial resistance**—to demonstrate interoperability so others can plug in the workflows of their choice.

Added value of the EOSC Federation

The EOSC Federation's chief value to the use case is its **lowering of the barrier for cross-border and cross-domain data sharing** for microbiologists and epidemiologists who lack heavy compute, storage or workflow-engineering expertise. With the Federation, a researcher could kick off complex, Europe-wide comparisons from a laptop, while the computational burden is distributed across a shared infrastructure, creating a **"Schengen" for data in Europe**.

Major challenges remain: the use of sensitive data demands robust and **trustworthy authentication / authorization protocols** that can enforce legal compliance, and requires trusted policies aligned with initiatives such as One Health and the ongoing policy discussions at the European Centre for Disease Prevention and Control (ECDC). EOSC's AAI is a promising technical building block. By demonstrating trustworthy, federated, cross-border, cross-domain analysis—without eroding data sovereignty—the **EOSC Federation will unlock faster responses to outbreaks and enable more ambitious, Europe-scale questions in pathogen research**.

