# SEARCH Federated Platform: Federated Data Discovery and Learning for Synthetic Data Generation in Healthcare

Chrysostomos Symvoulidis, Charalampos Geremtzes, Christos Ntais, Spyros Kollias

## FACTS & FIGURES

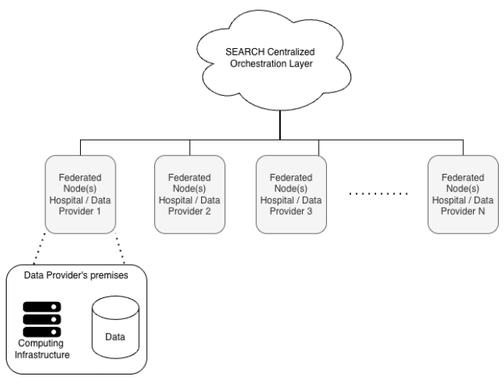
| | |
|---|---|
| Start date: | 1 October 2024 |
| End date: | 20 September 2028 |
| IHI funding: | 7.916.214,30 € |
| Industry contribution: | 2.753.005 € |
| Contributing partners: | 3.733.843 € |
| Total Cost: | 14.403.062,30 € |
| Project website: | ihi-search.eu |
| Social media: | LinkedIn |

## CHALLENGE

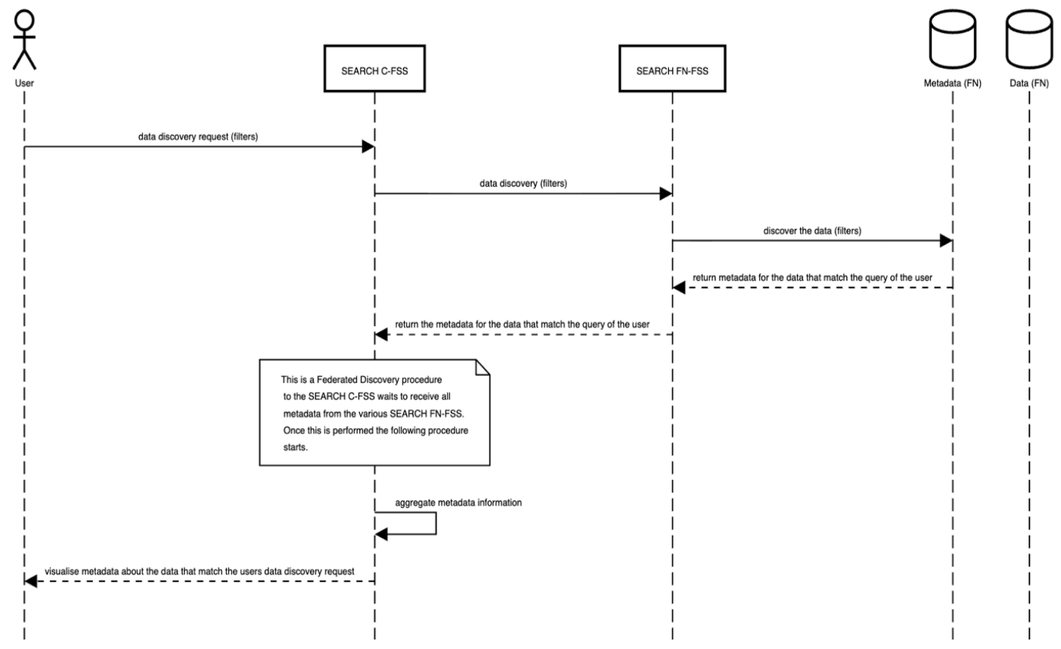
Healthcare AI requires access to large, diverse datasets, but privacy and GDPR constraints limit centralised data sharing. Traditional anonymisation can also reduce data utility.

SEARCH addresses this challenge through a privacy-preserving federated platform for synthetic data generation, enabling institutions to collaboratively train AI models while sensitive patient data remains local.

## APPROACH & METHODOLOGY




Data Discovery on the FSS Module

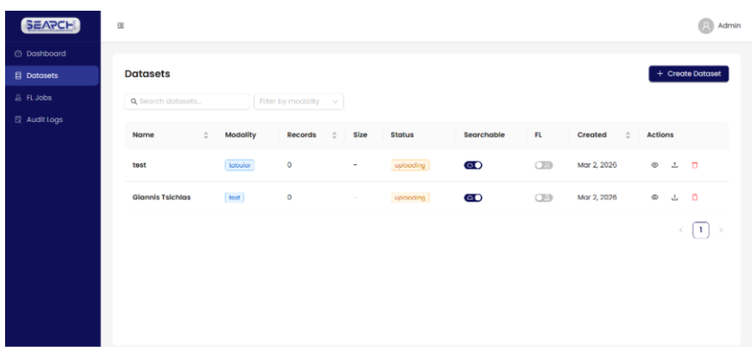


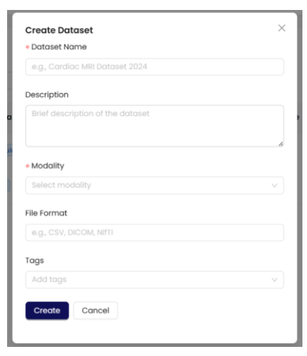

The SEARCH Federated Platform is designed as a federated, privacy-preserving ecosystem enabling the preparation, harmonization, and use of sensitive clinical data for synthetic data generation and AI model development. The platform is organised around two core workflows: (1) *collaborative federated training of SDG models* using harmonised datasets across data providers, and (2) the use of validated SDG models to *generate synthetic datasets* for downstream tasks, such as training diagnostic AI models. To achieve this, the architecture of the SEARCH Federated Platform, is comprised of the following modules.

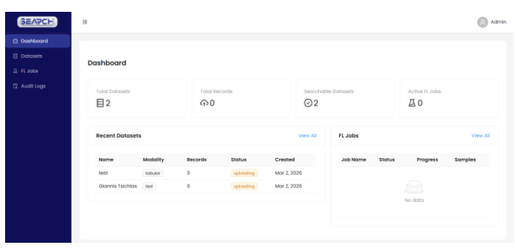
- The **Dataset and SDG Model Catalogue / Federated Storage and Search (FSS) Module** provides the core discovery, storage, and retrieval infrastructure of the SEARCH platform.
  - Data Storage: The actual data will be stored in an object storage (MinIO) object store. Elasticsearch will be used as the indexing database.
- The **Federated Learning Module (FLM)** enables collaborative model training without requiring any institution to share raw or harmonised data. Its core components are:
  - The *Global FL* deployed on the Centralized Orchestration Layer responsible for coordinating the federated learning process within the SEARCH Federated Platform.
  - The *Local FL* is a component deployed on the Federated Nodes layer and is responsible for performing the local training of the SDG model within the premises of each Data Provider enrolled with the SEARCH Federated Platform.

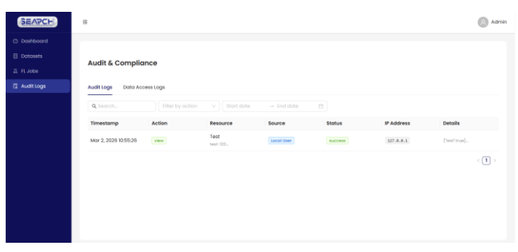
## RESULTS

The SEARCH Federated Platform provides an end-to-end workflow for privacy-preserving AI: discover distributed datasets, launch federated training through the Global Service APIs, monitor progress round by round, and download the resulting model. Raw data remains within the institution that owns it throughout the process.



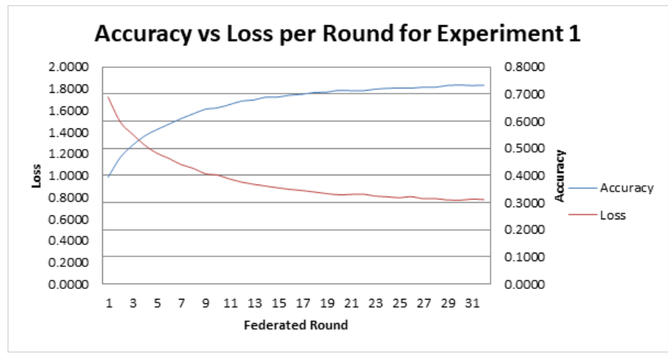




### CIFAR-10 CLASSIFIER

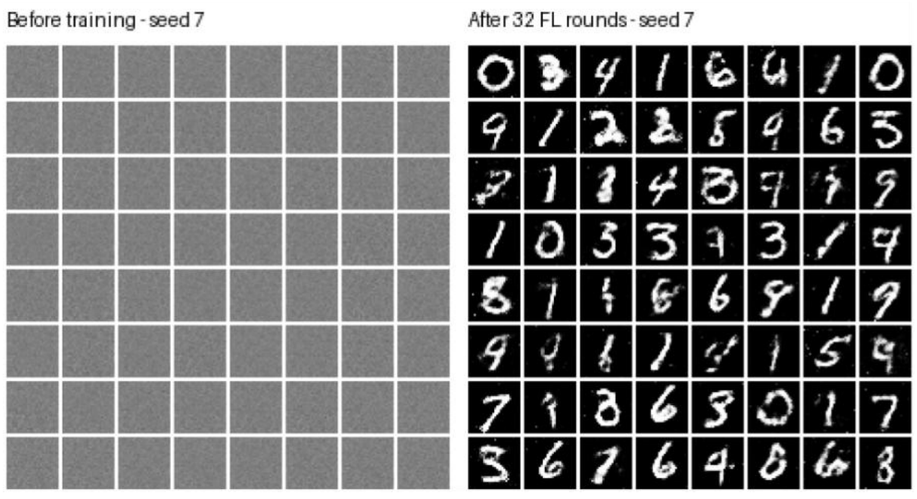
**0.39 → 0.73 global accuracy**

**32 rounds • 0 round failures**




### FEDERATED GAN • MNIST

**Recognisable synthetic digits after 32 federated rounds**




Both experiments completed the full lifecycle—from experiment creation to model download—through the platform APIs. Together, they validate the Federated Learning Module as the training layer connecting data discovery, harmonisation and model training.

Future work will focus on finalising the implementation of the SEARCH Federated Platform and evaluating it through the project's pilots, which will apply the platform to real data and real-world clinical use case scenarios.

## VALUE OF IHI COLLABORATION

IHI's public-private ecosystem enabled us to design the SEARCH Federated Platform around real healthcare infrastructure and governance constraints. Requirements from hospital and industry partners directly shaped architectural decisions, including Federated Nodes that do not expose inbound endpoints.

## CAREER IMPACT

Developing the SEARCH Federated Platform strengthened expertise in federated learning, distributed systems and privacy-preserving AI. Working with clinical and regulatory partners also provided practical experience in translating GDPR and healthcare requirements into technical architecture, broadening career interests towards privacy-preserving AI in healthcare.

Co-funded by the European Union

This project is supported by the Innovative Health Initiative Joint Undertaking and its members.