Federated Exploratory Data Analysis (FEDA)

Erik Ostrowski, Sumit Madan, Holger Fröhlich (Fraunhofer SCAI)

FACTS & FIGURES

- Start date: April 1, 2023
- End date: March 31, 2028
- IHI funding: 23,032,209 €
- Industry contribution: 19,671,182 €
- Total Cost: 42,703,391 €
- Project website: https://www.iderha.org/
- Social media: https://www.linkedin.com/company/iderha

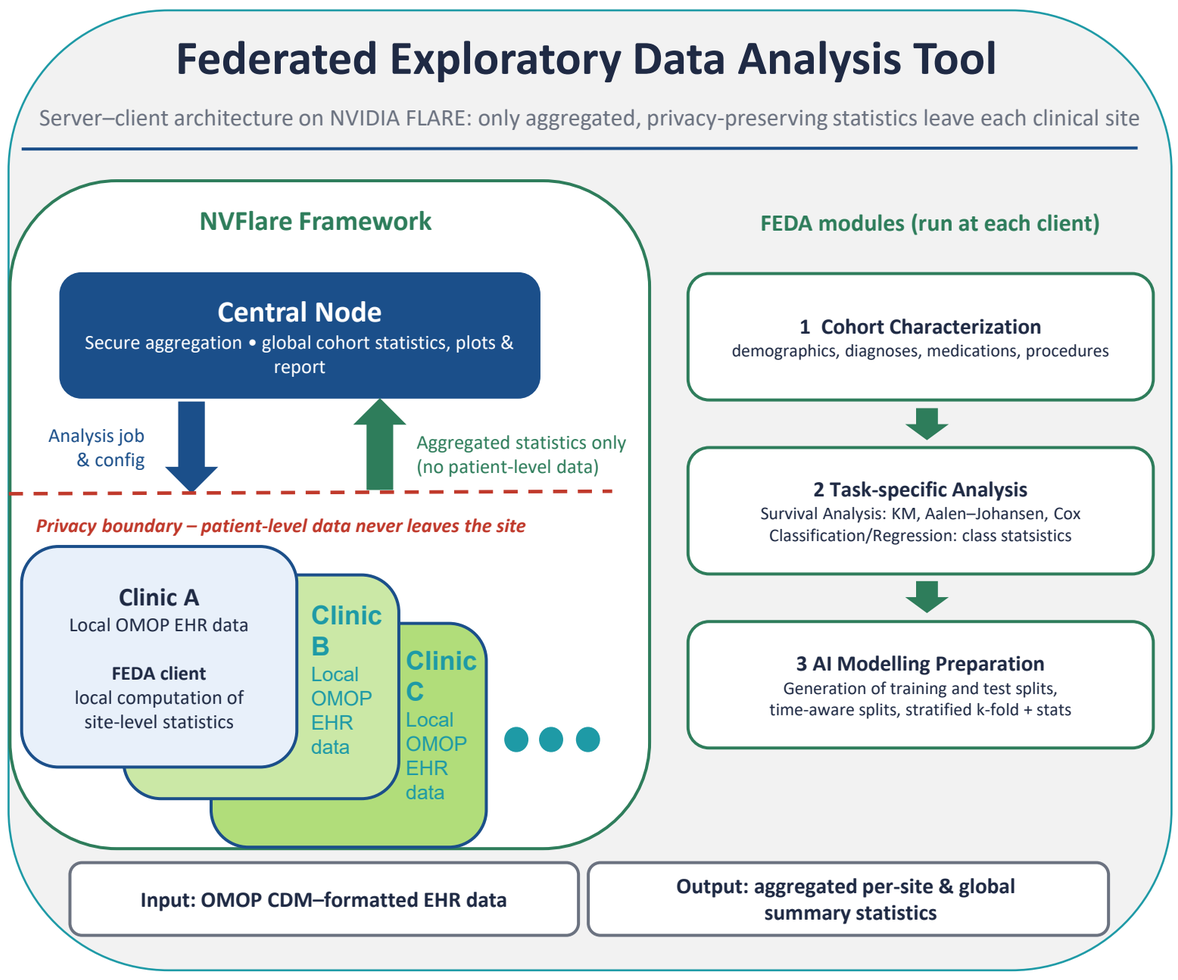
CHALLENGE

Federated Learning (FL) operates on data, which is not visible to data scientists. This is a challenge for algorithm development and testing.

- **Lung cancer** is Europe's deadliest cancer; the routine hospital data needed for early risk profiling exist but **cannot leave the clinic**.
- **Federated learning** moves model; the patient data stay put at each site.
- Blind spot: site **cohorts could vary** in demographics, coding, event rates, and missingness.
- No **global overview** of the combined sites as one dataset

The gap: No automated way to assess the cross-site cohorts and select appropriate models for federated learning.

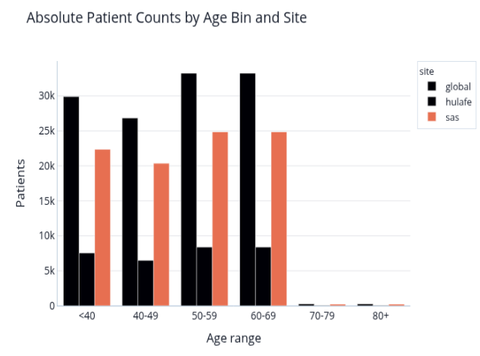
APPROACH & METHODOLOGY

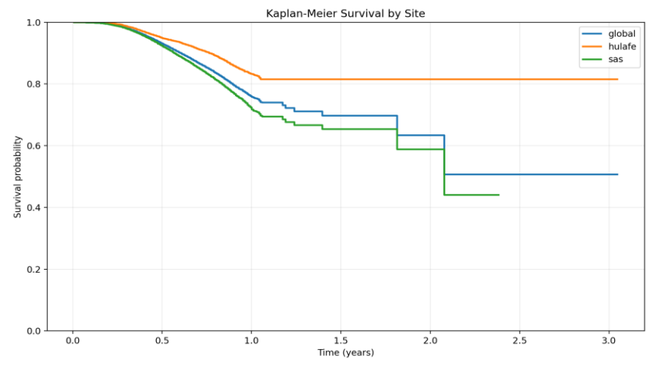


Our Federated Exploratory Data Analysis (FEDA) tool supports federated analytics for OMOP-based electronic health records (EHR). It uses NVFlare federated learning framework under its core. FEDA generates aggregated site-wise and global summary statistics reports.

RESULTS

- **Runs in the real world:** deployed across two Spanish clinical sites to assess lung caner risk prediction, ≥ 20,000 patients each
- **Complete cohort picture:** demographics, diagnoses, medications, procedures — aggregated per site and global statistics.
- **Federated survival analysis:** Plots such as Kaplan–Meier and Aalen–Johansen estimators
- **Open & reusable federated analytics:** modular, OMOP-native, on GitHub.





VALUE OF IHI COLLABORATION

- FL platform implementation **requires multidisciplinary and cross-sector expertise** in data science, IT, regulatory, legal, and clinical sciences. This can only be realized through a PPP.
- **Real hospitals, not simulations.** IHI's public–private setting enabled access to clinical institutions through federated learning infrastructure.
- **Leveraging open standards.** Built on OMOP Common Data Model, adopted by 140+ European organisations through the IHI-funded EHDEN project
- **Impact beyond one use case.** FEDA is disease and task-agnostic; reusable for federated OMOP analysis.
- **From research to long-term impact.** FEDA makes federated data legible for AI modelling. It accelerates model development by revealing data insights and supports clinical translational research.

CAREER IMPACT

- **Madan:** Younger scientists gain experience with leading international and interdisciplinary project teams
- **Ostrowski:** moving from classical centralized AI modelling to federated analytics and learning using NVIDIA FLARE orchestration, secure aggregation, multi-site deployment on physically separated servers, with real world lung cancer cohorts.
- **Ostrowski:** hands-on expertise in the OMOP CDM / OHDSI ecosystem.
- **Both:** coordinating successfully with clinicians, IT departments, legal experts and industry partners in three countries.

Co-funded by the European Union

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

Funded by the European Union, the private members, and those contributing partners of the IHI JU. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the aforementioned parties. Neither of the aforementioned parties can be held responsible for them.