

EPTRI General Assembly and Scientific Meeting

Warsaw, 28-29 July, 2026

Building a Federated Ecosystem for Paediatric Health Data

Rafał P. Bartczuk
Children's Memorial Health Institute

Nothing to disclose

Building a Federated Ecosystem for Paediatric Health Data

European paediatric institutions already generate large amounts of valuable data

These data come from:



Clinical care



Laboratory diagnostics



Imaging



Genomics



Biobanks



Clinical trials



Digital devices

The problem is not that data do not exist. The problem is that these data remain fragmented across institutions, structured differently, and difficult to find, search, and combine for research purposes.

❏ This challenge has been explicitly recognised in the context of the European Health Data Space, where paediatric-specific data needs were flagged as were flagged as insufficiently addressed in the current regulatory design.

Why Paediatrics Needs Federation

01

Small populations, rare conditions

~75% of rare diseases manifest in childhood, yet no yet no single centre sees enough comparable cases to cases to draw meaningful conclusions. Scale requires requires collaboration.

02

Children are not smaller adults

Age, maturation, and developmental stage all change how lab values, symptoms, and outcomes are interpreted. The same measurement can mean something entirely different across ages.

03

Research is inherently longitudinal

Paediatric research spans years — covering development, care transitions, and outcomes extending into adulthood. No single institution can can follow enough patients for long enough.

04

Highly multimodal data

Relevant data spans clinical observations, imaging, genomics, biosamples, patient-biosamples, patient- and parent-reported outcomes, and monitoring devices. Scale and standardisation are essential.

05

Additional safeguards required

Children's data demands parental consent, child assent, re-consent as children children mature, and long-term stewardship. Governance frameworks must be must be robust and age-sensitive.

❏ In paediatrics, no single hospital is likely to hold enough patients, enough observations, and enough types of data to answer every important research important research question.

The Case That Isn't Missing



A child is diagnosed and treated correctly. The clinical record is rich —imaging, genomics, lab results, years of follow-up.

Within one institution

One or two comparable cases. Not enough to study prognosis, prognosis, treatment response, or long-term outcomes.

Across European centres

Bologna, Barcelona, Munich, Warsaw —isolated cases become a cohort. Patterns emerge. Subgroups become visible.

Clinically, the patient is not missing. Analytically, they are almost invisible —
federation.

What is a federated data ecosystem?

Not a single model — but a range of increasing integration, each with different requirements and capabilities.

Level	What Is Shared	Real-World Example
1 — Discovery	Metadata only — what datasets exist, which variables they contain, contain, how data was collected. Underlying data never leave their original location.	European Open Science Cloud (EOSC) — interconnected nodes sharing catalogues
2 — Common Protocol Protocol	A shared analytical protocol is executed locally at each site. Only approved, aggregated outputs are returned for joint analysis.	DARWIN EU — distributed real-world data network network supporting EMA regulatory decisions
3 — Distributed Query	Standardised queries or statistical code are distributed to nodes built nodes built on a common data model.	Networks using OMOP common data model
4 — Federated Learning Learning	Algorithms trained locally at each site. Only model parameters or updates — never raw data — are exchanged and combined centrally.	Machine learning models trained on distributed distributed clinical data

The architecture differs across levels. The principle is constant: local responsibility, collective learning.

Federation Is Not Centralisation

Centralised Model

Data is copied to a single environment for analysis.

- One repository
- Central governance
- Extensive data transfers
- Easier technical uniformity

Federated Model

Data remains at participating nodes, with coordinated execution.

- Local responsibility
- Shared standards
- Distributed execution
- Coordinated governance

❏ **Federation does not remove complexity.** It redistributes it—preserving autonomy while enabling collaboration.

Federation in Practice: European Initiatives

Federation is not just a future concept; several European initiatives are already developing and implementing different aspects of this model.

1

European Open Science Cloud (EOSC)

A production-grade federation of national, thematic, and infrastructure nodes, connecting scientific data, services, and computing resources across research communities. As of July 2026, EOSC included the EU Node and 27 candidate nodes.

2

DARWIN EU

An operational example in health research, connecting data partners that transform local observational health data into a common structure for distributed studies requested by European medicines regulators. By June 2026, DARWIN EU had onboarded 40 data partners.

3

PHEMS

A Pediatric Health Data Space specifically focusing on privacy-preserving use of data from European children's hospitals within ECHO, including decentralised analyses, like benchmarks.

These initiatives, while distinct, collectively demonstrate a clear direction: moving from isolated databases towards trusted networks of interoperable nodes.





PeRCMC: A Local Node for Federated Paediatrics

At the Children's Memorial Health Institute, we are developing the Paediatric Regional Centre for Digital Medicine (PeRCMC). Its purpose is to create the local capabilities the local capabilities required for both internal research and participation in larger data networks.



Integrated Data

Routine clinical data, lab results, imaging, pathology, pathology, genomic and other omic data, linked to to biobank samples.



Multidisciplinary Team

Combining clinical knowledge, data engineering, data management, biostatistics, data science, legal expertise, and information security.



Node Capabilities

Evaluating data quality, defining cohorts, pseudonymisation, access management, and executing analyses in a controlled environment.

PeRCMC should be understood not as a local data warehouse, but as an institutional capacity to participate responsibly in federated research.

IPCZD Real-World Data & OMOP CDM Status

Data source

- Hospital discharge, inpatient and outpatient specialist records (tertiary care)
- Established: 2016 | Data collection from: 2008
- Country: Poland | Language: Polish
- Total population: 561,392 | Active population: 59,461
- Paediatric focus (<18 years) incl. neonates & rare diseases

Governance & refresh

- Daily data capture from EHR
- Quarterly refresh | Last refresh: Dec 2025
- Informed consent required for all studies
- Data validation against source medical charts possible

Data content

- Diagnoses (ICD-9)
- Procedures (ICD-9-CM, SNOMED CT)
- Measurements (LOINC / SNOMED CT)
- Prescriptions & dispensing (RxNorm)
- ICU admission, cause of death, healthcare utilisation units
- Unique patient identifier | Permanent family linkage (mother–child, father–child, siblings)

CDM status

- OMOP CDM implemented
- ETL status: In progress

Key Lessons from Preparing PeRCMC for FE

Our journey in developing PeRCMC and preparing for participation in federated networks like DARWIN EU and PHDS has yielded crucial insights: First of all **data harmonization is the first and important challenge**, second is **regulatory harmonization**.

01 harmonization.

01

Start with Real Use Case Helps

Data harmonisation is best driven by specific research questions, which clarify essential data elements, definitions, and quality thresholds.

03

Measure Data Quality Actively

Data quality metrics (completeness, consistency, plausibility, temporal relationships, provenance) must be evaluated systematically, systematically, not assumed.

02

Clinical Interpretation for Data Mapping

Effective data mapping goes beyond technical process, requiring deep requiring deep clinical understanding of how local codes are generated and used.

04

Local Teams Are Essential

Federated ecosystems depend on empowered local teams and and distributed expertise, avoiding concentration in a central coordination office.



Three Key Messages



Fragmented Paediatric Data

Isolated data is often too small, heterogeneous, heterogeneous, and dispersed to yield its full full scientific value.



Federation: Not Centralisation

It means shared definitions, governance, and and procedures, preserving local data responsibility, not a single database.



EPTRI's Connecting Role

EPTRI can provide a trusted European layer of standards, services, expertise, and clinical validation for paediatric research.

Our objective is not to move every dataset, but to empower institutions to collectively answer vital paediatric questions. By connecting trusted nodes, we can make nodes, we can make analytically missing cases visible, creating stronger evidence for children across Europe.