the hyve

FAIR Biomedical Data Services and Solutions

For top 20 pharma companies and academic partners

Rosa Bianca Gallo, Sales Executive rosa@thehyve.nl

Index

the hyve

We advocate for open sciences and open source software for research

Data Ops team












https://thehyve.nl/about-us/team/

Skills

Data Engineering

- Standard Biomedical Data Models (OMOP, CDISC, ADAM, SDTM, FHIR, i2b2, etc.)
  - Customization / extension of data models, vocabularies and concepts

- Deployment of helper tools such as data mapping tools & ontology management tools
- Semantic modelling & RDF graphs / triple store DBs

Domain Expertise

- Biomedical informatics expertise spanning entire biopharma R&D and drug development lifecycle
- Computational biology including genetics & genomics, biomarker discovery
  - Translational medicine, integration of clinical, biobanking, imaging and genomics data
  - Real world data, epidemiology and digital biomarkers

Other Scientific Domains

- Epigenetics, systems biology, plant and mammalian bioinformatics, immunoinformatics
- Biomarker discovery, genome assemblies and annotation
- High performance computing (HPC) and experience with public resources like NIH, GEO, etc.

Open Source Software

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

RADAR-BASE

cBioPortal
FOR CANCER GENOMICS

Open Targets

Fairspace

Precompetitive Health Data Projects

EHDEN
EUROPEAN HEALTH DATA & EVIDENCE NETWORK

FAIRplus

BigData@Heart
Big Data for Better Hearts

EMIF

CINECA

PIONEER

RADAR-AD

trait
a Lygature platform

FNS - Cloud
Food Nutrition Security

RADAR-CNS

QuaLiFY

eTRIKS

ND4BB

TRANSLOCATION

eboVAC

National Health Data Networks

SPHN
Swiss Personalized Health Network

health RI
research infrastructure

Genomics england

MEDICAL INFORMATICS INITIATIVE GERMANY

Partner Communities

Pistoia Alliance

elixir

DTL
DUTCH TECHCENTRE FOR LIFE SCIENCES

GO FAIR

imi
innovative medicines initiative

Global Alliance
for Genomics & Health

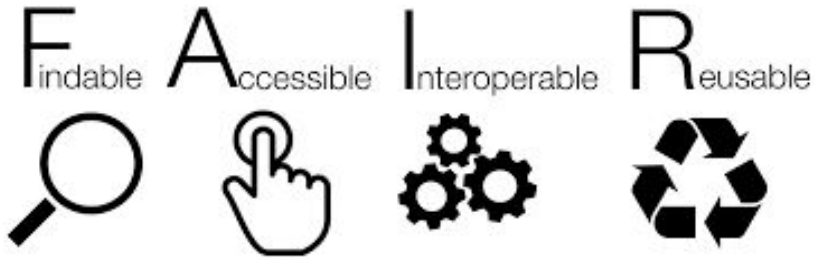
IT and Data Science services

IT Services

- Software Development
- Hosting & Maintenance
- Training

Data Services

- Data FAIRification
- Semantic modelling
- Knowledge graphs



Consulting Services

Top down and Bottom up approaches are supported by our consulting team with FAIR frameworks to help our clients with:

- *FAIR assessment of data assets, processes and tools*
- *Gap analysis for Information Architecture*
- *Customization of (meta-)data models and standards*
- *Data Management and Governance Plans and schemes*
- *BYOD workshops*

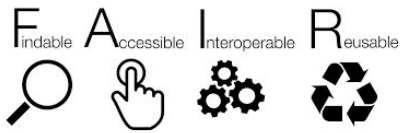
Type of Projects

- Archeology / unveiling valuable data
- Cataloguing / Landscaping
- Metadata definition and development
- Harmonization / curation projects
- Knowledge Graphs design and setup
- Power sessions to reach consensus on data models, schemas, metrics, etc.

Commercial Use cases

1. FAIR assessment of R&D informatics tool
2. FAIR data-thon
3. FAIR blueprint for Value Based Health Care
4. Conformance of Clinical Trials data
5. Semantic Model for Functional Genomics
6. Data Landscape of R&D
7. Metadata Convention
8. Ontologies mapping and enrichment
9. Cohort selection for longitudinal data
10. RWD for Pediatric Virtual Clinical Trials
11. RWE generation from large scale population data
12. Support Data Office with Data Governance Tooling
13. Linked EHR data
14. Standardization of commercial data assets to OMOP CDM (Truven, Optum, Symphony, CPRD, etc)
15. RWD from sensors - data harmonization

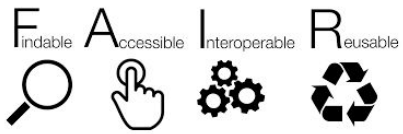
1. FAIR assessment



We have been contracted as an external (FAIR) advisor to assess the FAIRness of a tool built in house by the R&D informatics team (top 10).

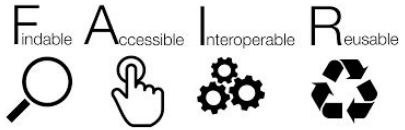
We also moderated a datathon to help strengthen the "cultural" layer of FAIR.

2. FAIR data-thon



For this top 10 pharma client we moderated a datathon to help strengthen the "cultural" layer of FAIR with a BYOD series of sessions, supporting change management and digital transformation leaders in the identification of scientific use cases that will foster a collaborative mindset.

3. FAIR blueprint for evidence based health care



For this top 10 pharma client, we consulted with the CDO and Head of Data Sustainability around the best data models and data architecture for a Value Based Healthcare Platform.

4. Conformance of Clinical Trial Data

For a corporate program aimed at the digital transformation of the organization, we created a ***"conformance layer"*** *that allows for a semantic representation of clinical trial and observational data to benefit search and analysis applications.*

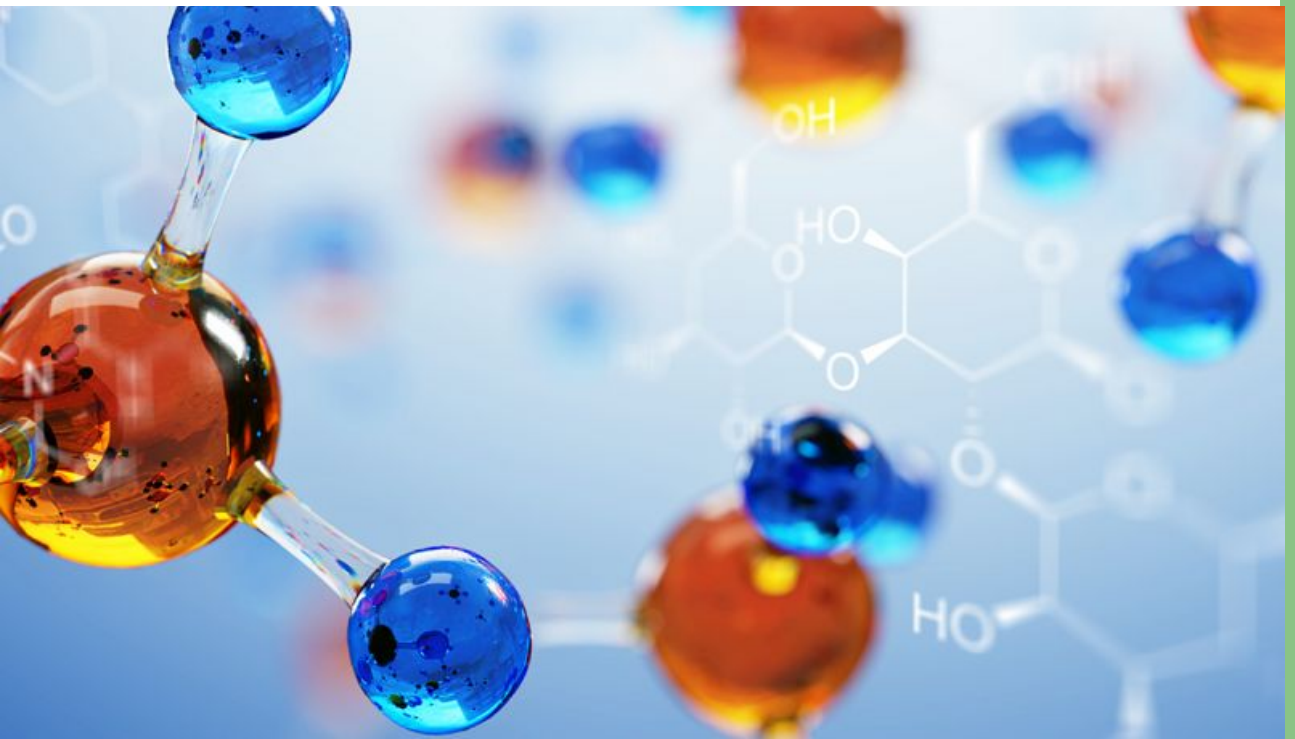
Such an integrated data environment supports cross-domain analytics activities.

5. Annotation of Genomics Data

For a large corporate program we created a semantic model for conformance of clinical trials. In this use case we describe our work done to extend that model for the Computational Genomics group within a so called silo-breaker effort.

We developed a genomics model to enable easier metadata annotation of datasets. Ultimately, this will facilitate ingestion into cBioPortal on an ongoing basis.

6. R&D Data Landscape



For this top 10 pharma client we created a knowledge graph to enable search on (meta)data spanning from early-stage research to late-stage clinical trials.

The knowledge graph leveraged a customized semantic model of the applications and dependencies in the data landscape.

7. Metadata Conventions

For this top 10 pharma client, we have developed a pragmatic proposal for standardization of the most widely used metadata elements. Deliverables:

- *Metadata inventorisation; from data catalog and metadata registry*
- *Metadata assessment*
- *Developing conventions for URIs minting (defining URI pattern)*
- *Aligning metadata with enterprise data governance (existing open ontologies)*

8. Ontology Mapping

For this top 10 pharma client we assist with annotation of the custom metadata descriptions to a series of public ontologies to support richer querying.

This allows the client to harmonise data across experiments to drive advanced searches and improve data visualisation.

9. Cohort selection & exploratory analysis

Glowing Bear

Glowing Bear is the user interface we built on top of i2b2/tranSMART data model. It provides support for queries on longitudinal and sample data

- *Cross-study, ontology-based variable harmonisation*
- *Decoupled user interface, built on Angular*
- *API-level enforced security*
- *Standardized installation with Docker*

10. Real World Big Data streams for remote Clinical Trials

RADAR-BASE

For this top EU based CRO, we adapted the open source platform, RADAR-base, to enable sensor data collection and management of large scale unstructured datasets in the context of remote clinical trials for pediatric patients with respiratory diseases and depression.

11. RWE generation from large scale population data

For this top EU based pharma, we adapted an Open Source platform, OHDSI, and in particular the Analytics component of the suite, called Atlas, to enable a per-patient view and custom cohort selections over large scale population data.

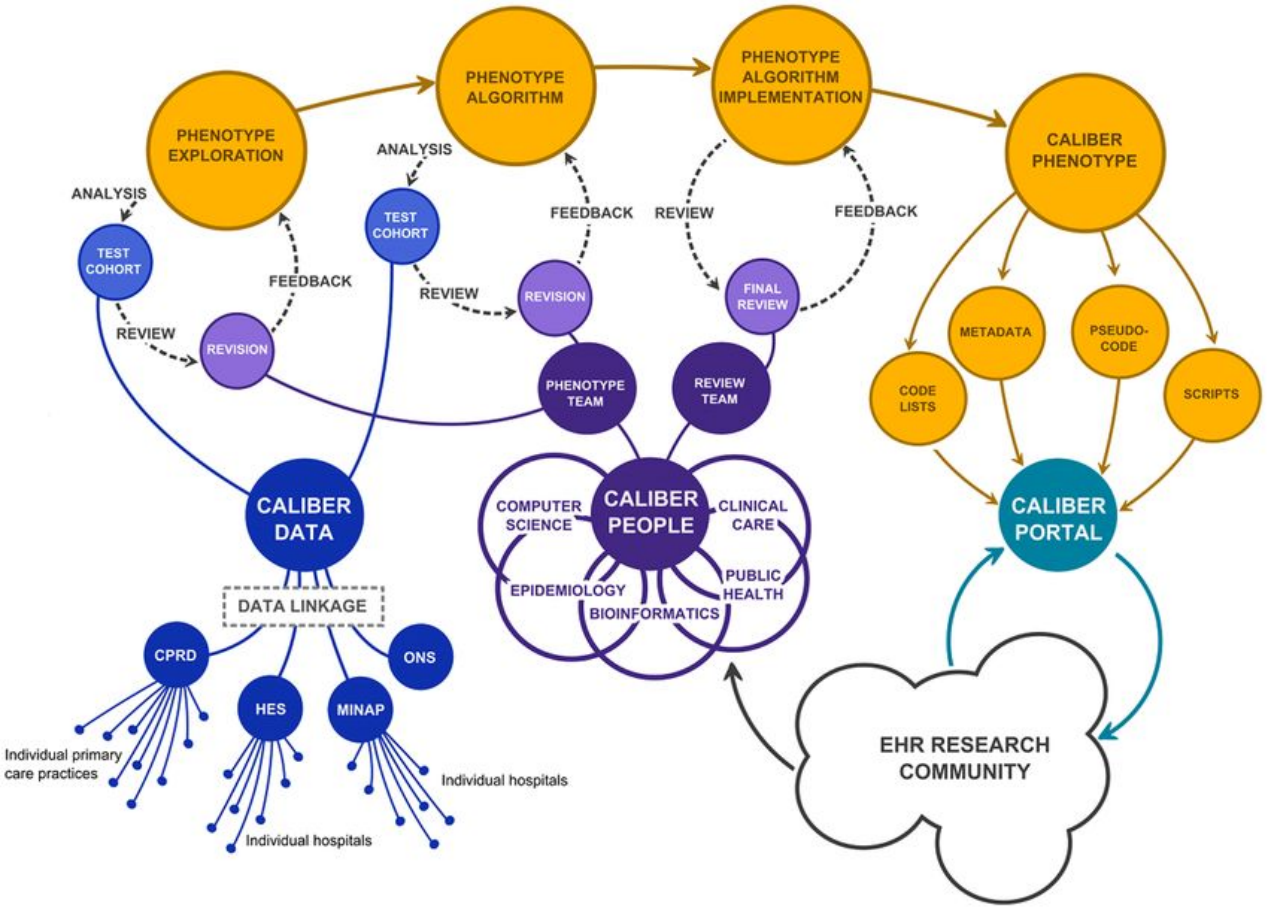
Read more here.

12. Support Data Office with tooling for data governance

Fairspace

For this large European academic hospital, we created a platform that can be used by data stewards to design data models, and by researchers to annotate and publish their existing research data as digital FAIR data objects.

13.Linked EHR data, CALIBER



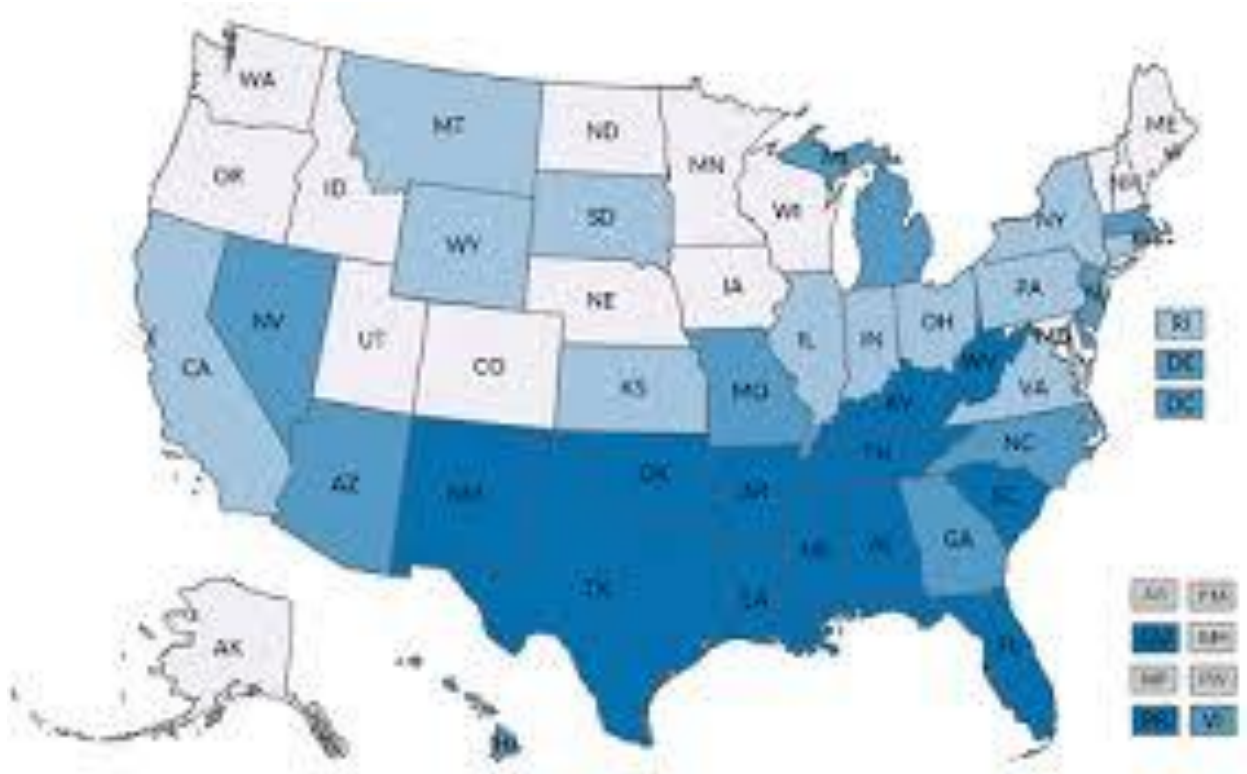
For UCL health, in the context of BigData@Heart. we mapped the Caliber study data to the OMOP CDM. Caliber is a UK phenomics platform for developing and validating EHR phenotypes. Translational research platform
linking national structured data
and socioeconomic information
from

- *primary care (CPRD)*
- *hospital care (HES)*
- *mortality registry (ONS)*

We helped linking patient level data from 3 sources: the Clinical Practice Research Datalink, Hospital Episode Statistics and the Office for National Statistics) to OMOP.

[Source]

14.Truven Health MarketScan Database for Research OMOP Conversion



For Top 10 Pharma clients we map MarketScan data, alongside other data types, to the OMOP CDM and harmonize with different data sources to enable successful analysis.

We also worked with Optum, CPRD and other licensed types of data.

15.Wearable Devices Data Conversion

For this global non profit foundation, pioneering the research around Parkinson's disease, we harmonized datasets from different wearable devices, including smartphone data, smartwatch (Motorola) data, Empathica, Physilog and integrated with diary data.

Public-Private Consortia

Five-years project aiming to ensure the optimal care for all European men living with prostate cancer by unlocking the potential of Big Data and Analytics. The project is still in its early stages and The Hyve is collaborating with E.F.P.I.A. European Federation of Pharmaceutical Industries partners and academia to set up a catalogue of relevant prostate cancer registries and other useful healthcare datasets. The Hyve will use the **OMOP and OHDSI** technologies to support data integration and analysis for longitudinal prostate cancer registries, whereas **tranSMART** technology will be used for cohort studies with deep phenotypic and clinical data which also include omics data.

BigData@Heart is a five-years project aimed at setting standards for cardiovascular big-data science. Among other tasks, The Hyve is working on a case study which compares the survival of heart failure patients across England, Sweden, Spain and the Netherlands. The Hyve is currently in the process of converting health records from these four countries to the OMOP Common Data Model. This conversion allows us to leverage existing observational methods that are commonly used within the OHDSI community.

The combination of **tranSMART** and **OHDSI** has been used successfully in IMI EMIF. The European Medical Information Framework (EMIF) project is an IMI project, aiming to develop a common information framework of patient-level data that will link up and facilitate access to diverse medical and research data sources. Thereby opening up new ways of research for scientists. Obesity and dementia are two of the greatest healthcare challenges of our time. To guide the development of the framework, the team will initially focus on two key research issues:

- identifying the mechanisms that make some people more susceptible to dementias (such as Alzheimer's disease) than others
- determining which individuals with obesity are most likely to develop complications such as diabetes.

The follow-up project of EMIF, the **European Health Data and Evidence Network** (EHDEN), has just launched in November 2018. This network aims high: it wants to create a common standard for hundreds of health data silos all over Europe. This standardization should boost medical research and healthcare outcomes, and especially support the BD4BO consortia. The Hyve plays a key role in the consortium by **leading WP4 (Technical Implementation)**.

innovative medicines initiative

FAIRplus

FAIRplus is a precompetitive initiative with 22 participants (12 academic, 7 EFPIA, 3 SMEs)that aims to develop tools and guidelines for making life science data FAIR (Findable, Accessible, Interoperable, Reusable). The project has 22 partners from academia and industry, and runs from January 2019 to June 2022. See About the project and the news story on the ELIXIR website.

The Hyve holds coordinating role in one of the WPs and a leading role in the Squads. Technical contribution of The Hyve involves:

1. FAIRification of IMI datasets (mainly omics data)
2. Creating data management plans
3. Contributing to FAIRPlus cookbook recipes; current contribution: how to request new terms in public ontologies

the hyve

We empower scientists by building
on open source software