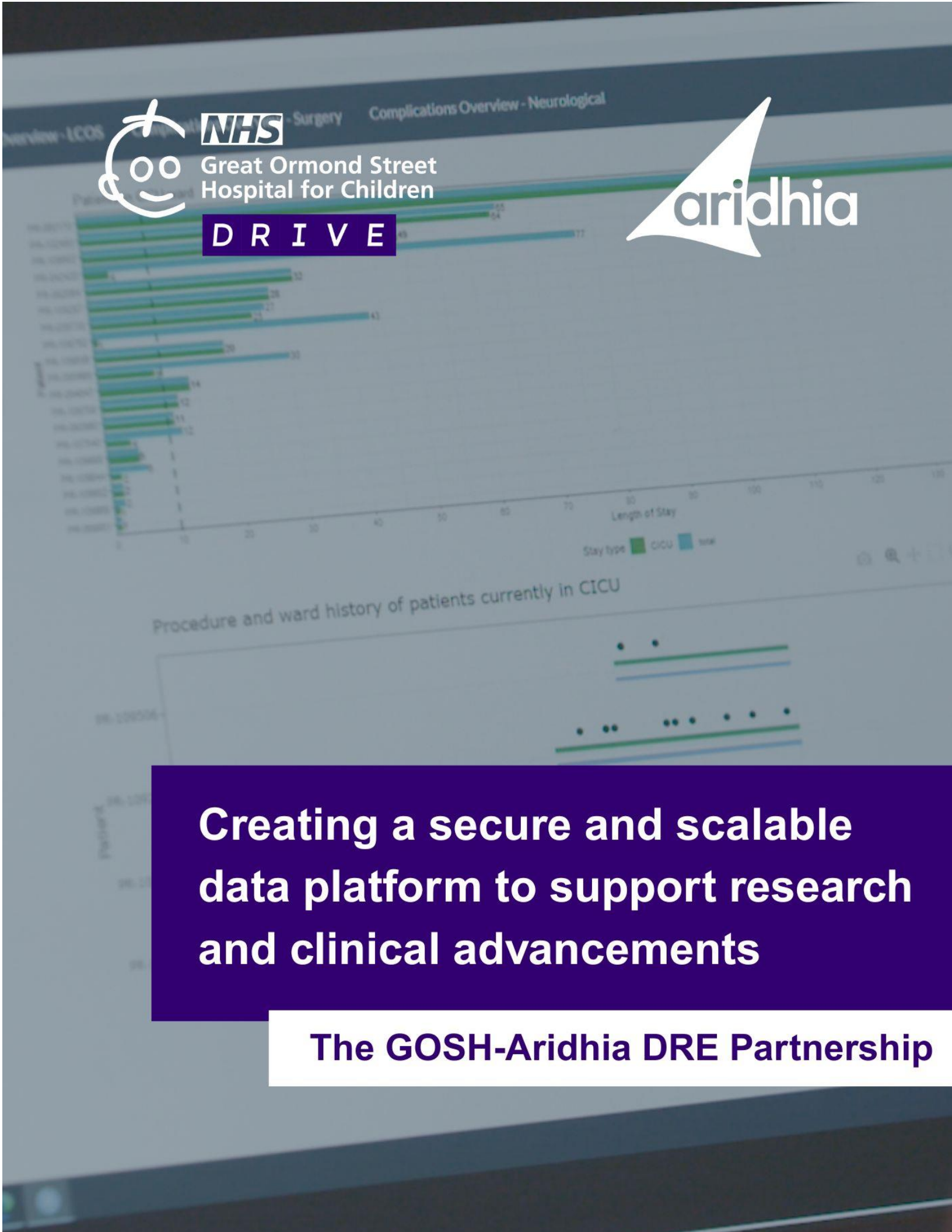


Creating a secure and scalable data platform to support research and clinical advancements

The GOSH-Aridhia DRE Partnership

Contents

Alicia Gibson[2], Neil Sebire[1], Daniel Key,[1] Lydia Briggs[1], John Booth[1], Ross Stiven[2], Kim Carter[2], Robert Bryce[2] , Andrew Taylor[1], Shiren Patel[1], David Sibbald[2]

[1]GOSH, [2]Aridhia - Published April 2025

GOSH DRE Development Over the Last 8 Years:

Great Ormond Street Hospital for Children NHS Foundation Trust (GOSH) is one of the most digitally advanced hospitals in Europe. The implementation of their comprehensive electronic patient record paired with a state-of-the-art secure, trusted digital research environment sets them apart as an Intelligent Research Hospital. They are a go-to centre for innovation and aim to leverage data and digital technologies to improve patient care and staff experience.

Work began on the trusted Digital Research Environment (DRE) back in 2017/8 with the establishment of the GOSH Data Research, Innovation and Virtual Environment (DRIVE) unit as a centre of engagement for paediatric innovation.

In 2017 GOSH partnered with Aridhia Informatics to build a state-of-the-art DRE infrastructure. In parallel GOSH created a physical space, DRIVE, and began to build a team of digital engineers, governance experts, data scientists and clinical experts.

The overall aim was to harness the secondary use of routinely collected data from EPR and other systems to improve the operational management and planning of the hospital and to support clinical and informatics research activity.

GOSH entered a ten-year strategic partnership with Aridhia to support the research and innovation component of the Trust's multi-year digital transformation programme. This initiative was designed to establish a secure, scalable research platform that enabled clinical and research teams to collaborate effectively, conduct advanced data analysis, and generate rapid, auditable, and reproducible research outcomes.

By leveraging Aridhia's expertise in data-driven research environments, GOSH's aim was to enhance its capabilities in biomedical research, precision medicine, and clinical innovation, ensuring that healthcare advancements are underpinned by a robust, compliant, and future-proof digital infrastructure.

The platform was integrated into the Trust's new EPIC Electronic Patient Record (EPR) system and revamped the way in which GOSH managed its research projects.

Up until this point, GOSH found itself in a technical situation that did not match its reputation as a world leader in paediatric health research. The primary aim was to speed up delivery and extraction of data for research: not to produce the best data warehouse. The secondary aim was to be the single source of truth for data for research, and finally, to provide consistency of delivery to all projects.

During this time, the DRE Team at GOSH grew significantly, including governance specialists, data engineers, data scientists and clinical experts.

Infrastructure Growth and Changes

In its first 8 years, the DRE has experienced remarkable growth, evolving to hosting over 400 projects. This transformation highlights the platform's growing role in enabling data-driven research and innovation. Initially hosted on the private cloud provider UK Cloud, the DRE underwent a significant transition to Microsoft Azure during 2019/2020. This move followed the NHS's approval of public cloud platforms for storing patient data and was executed through a strong collaborative partnership between GOSH and Aridhia.

The migration to Azure was a pivotal moment in the DRE's evolution, enabling enhanced scalability and performance. Since transitioning to Azure, the DRE has expanded significantly, growing from a single UK hub to a multi-hub architecture that now includes three distinct hubs. Among these is a specialist Hospital Workspaces Hub, designed with enhanced security features to address the specific needs of sensitive healthcare environments.

A further milestone was the introduction of a centralized DRE Landing Portal, which provides seamless access to the DRE's key services, including FAIR Data Services and multiple Workspace hubs. This was followed by Single Sign On (SSO) integration for seamless authentication.

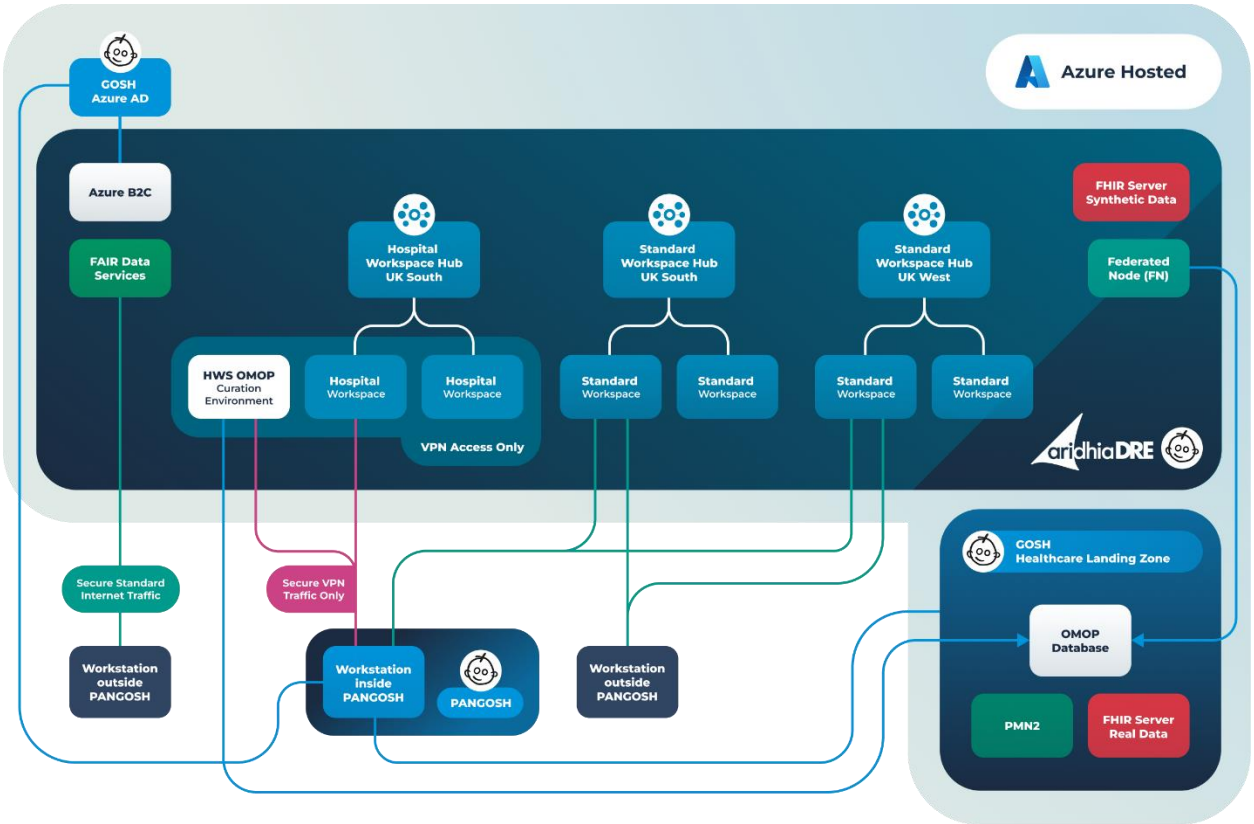


Figure 1: The DRE has not just grown in scale but also in complexity.

Prior to 2020, data management features within the GOSH DRE, such as cohort selection and data ontology, were dispersed across multiple standalone tools, each requiring separate logins. In 2021, with input from the GOSH DRE team these fragmented tools were integrated into a single cohesive platform within the Aridhia FAIR Data Service product.

This integration has significantly enhanced the user experience and functionality of the DRE. Aridhia FAIR Data Services now serves as one of the two core pillars of the DRE, empowering researchers and innovators to:

- Discover and Understand Data
- Explore Metadata
- Streamline Data Access Requests (DAR).

The DRE's transformation underscores the importance of adaptability and collaboration in supporting data-driven healthcare advancements. By leveraging the scalability and security of Microsoft Azure, introducing centralised access through the landing portal, and integrating core features into the FAIR product, GOSH and Aridhia have created a platform that empowers authorised researchers to unlock the potential of healthcare data.

To further support scalable, cost-efficient AI-driven research, ongoing efforts are focused on implementing Zero-Copy Data Sharing, allowing datasets to be accessed without unnecessary duplication. Additionally, the adoption of columnar storage formats such as Parquet and Delta Lake will further optimize performance for AI-powered analytics and large-scale research workloads.

This journey not only highlights the success of the DRE to date but also lays a strong foundation for future innovation in health data research and federated data collaborations.

Security and Compliance

Ensuring data security and regulatory compliance is paramount to GOSH's research and clinical operations, shaping the development and implementation of the Aridhia DRE. Through close collaboration between the GOSH and Aridhia Information Security teams, the DRE has been designed to meet stringent security and governance requirements. The DRE operates within a highly secure and controlled framework, aligning with internationally recognised security and privacy standards, including ISO 27001 for information security management, ISO 27701 for privacy information management, and Cyber Essentials Plus certification. Additionally, it complies with NHS Digital's Data Security and Protection Toolkit (DSPT) and GDPR regulations, ensuring robust governance and accountability in data handling. The platform integrates advanced security measures such as role-based access controls, end-to-end encryption, audit logging, and identity management via trusted Identity Providers (IdPs) allowing for SSO. These stringent protocols provide GOSH with a reliable environment for securely managing and sharing sensitive patient and research data while facilitating seamless collaboration across multidisciplinary teams and

institutions. Although Aridhia hosts the GOSH DRE, it does not have access to any data stored within the Workspaces.

Data Table Analytics Modules

One of the foundational collaborations between Aridhia and GOSH focused on addressing the need for intuitive, user-friendly tools to analyse and visualize data within the DRE. Recognizing the importance of empowering research teams with accessible analytics capabilities, the partnership led to the creation and integration of Data Table Analytics (DTA) modules into DRE Workspaces.

GOSH research project teams expressed a need for tools that would enable them to perform sophisticated statistical analyses and data visualizations without requiring advanced coding expertise. GOSH's vision was a point-and-click interface that would simplify data exploration and provide rapid insights. These tools would enable researchers to quickly visualize and analyse data, fostering a deeper understanding of complex datasets. By lowering technical barriers, the DTA modules would democratize access to statistical analysis, allowing teams with varying levels of technical expertise to actively engage in data-driven research.

In response, Aridhia and GOSH worked collaboratively to design and implement a suite of over 20 DTA modules into DRE Workspaces . These modules, embedded directly into DRE Workspaces, were developed to meet diverse analytical needs across a range of research applications. The modules include:

- **Statistical Modeling and Testing**: Tools for conducting advanced statistical analyses, including hypothesis testing and regression modeling.
- **Descriptive Statistics**: Modules for summarizing datasets and exploring key metrics such as distributions, means, and standard deviations.
- **Data Visualization**: Intuitive tools for creating charts, graphs, and plots to effectively communicate findings and trends.

To enhance usability further, the introduction of Aira, an AI-powered research assistant, will provide researchers with low-code/no-code AI capabilities, enabling them to derive insights from complex datasets more efficiently. These AI-driven tools will complement existing analytics modules by offering natural language query support, automated data summarization, and AI-assisted workflow automation within research environments.

Data Types and Volumes

Over the first 8 years, the Aridhia-supplied DRE has enabled GOSH to carry out pioneering data projects with a range of collaborators and industry partners. The project data has become multi-modal and grown in size and complexity.

Case Study: Genomics Information Finder

At GOSH a significant portion of genomic data is stored as PDF documents within the Electronic Patient Record (EPR) system, posing challenges for data accessibility, interoperability, and analysis. To address this, a Natural Language Processing (NLP) pipeline was developed in partnership with Roche UK, to extract and standardize genomic information, transforming it into a structured format within the Aridhia HLZ FHIR server. This approach established a single source of truth for genomic data in a standard format. This data could then be flattened into table representations for individual project cohorts in DRE workspaces. This was the process followed for multiple types of genetic data, e.g. germ-line Single Nucleotide Polymorphisms in rare disease and somatic gene fusions in our oncology. https://ebooks.iospress.nl/doi/10.3233/FAIA241059

Case Study: Processing Imaging Data

GOSH is currently developing a pipeline to integrate imaging data from the GOSH Vendor Neutral Archive (VNA) system into DRE workspaces. This process involves deploying an on-premises instance of the XNAT image management system to facilitate initial cohort extraction and de-identification. Additionally, an instance of XNAT will be deployed within the DRE to manage and process imaging data for specific research studies.

Case Study: Data Scaling

As the number of research projects at GOSH has increased, the volume of data has scaled accordingly. This growth has been seamlessly accommodated through the flexible cloud architecture of the DRE, which enables the effortless expansion of both individual workspaces and the overall number of available workspaces. This scalability ensures that researchers can continue to collaborate effectively without infrastructure limitations, supporting the hospital's commitment to advancing data-driven innovation.

Case Study: Data Pseudonymisation

Data pseudonymisation is a critical step in preparing data for analysis within research workspaces, ensuring compliance with privacy and security requirements. GOSH's data upload process integrates a pseudonymisation tool, PMN2, which is hosted within the GOSH's DRE. PMN2, an Aridhia-developed solution, enables the application of up to 16 configurable pseudonymisation rules. For example, the "Age Group" rule transforms specific numerical values or dates—such as age or date of birth—into predefined age ranges, converting an age of 7 into the category "0-9." This systematic approach enhances data privacy while maintaining the utility of the dataset for research and analysis.

Legacy Data & RDVs - How the GOSH DRE developed before EPIC.

The primary objective was to accelerate the delivery and extraction of data for research purposes, rather than focusing on creating the best data warehouse. The second goal was to establish a

single, reliable source of truth for research data and finally to ensure consistent data delivery across all projects.

Initial systems of interest:

- PIMS: Patient information and management.
    - In use from 2000.
    - Patients
    - Hospital admissions
    - Episodes of care
    - Ward stays
    - Outpatient visits
    - Coded diagnoses (ICD-10)
    - Coded procedures (OPCS-4)
- OMNI: Laboratory testing including microbiology & virology.
    - In use from 2000.
    - Lab specimen collection
- JAC: Electronic prescribing.
    - Started in 2006, complete hospital by 2010 except ICU.
    - not used in ICU, see carevue below.
    - medication orders
    - medication administrations
- Carevue: EPR system for ICU.
    - Fully active from 2013/14
    - medication orders and administrations
    - measurements
    - lab results, imported from OMNI.
    - clinical notes
- onBase: patient correspondence.
    - Individual documents held as PDF.
    - Backscan documents single PDF but held a number of scanned documents.

From the outset, the GOSH DRE team brought experience in extracting data from these existing systems, although previously this was done for specific purposes.

They initially accepted 10 proof-of-concept projects to evaluate the data requirements and understand how the data would be utilized.

A key decision was made not to provide a customized solution for each project. Instead, they opted to develop a composite solution that could be tailored to each project's needs. This solution would include the required data but would require each project to perform some data wrangling to format it as needed. It was believed that each project team should possess these skills, though the DRE team could offer training in basic R techniques for tasks such as combining datasets, grouping, and extraction. Providing immediate solutions addresses short-term needs, but equipping individuals

with the knowledge and tools to solve challenges independently ensures long-term sustainability and success.

At the time of this project, Workspaces only supported R with a very simple R Console. Workspaces now have RStudio built in.

They considered several existing models for data extraction, including FHIR, OpenEHR, and OMOP. Given their limited resources, it was determined that fully committing to any of these detailed models would be too time-consuming. As a result, the DRE team opted to develop their own simplified output structure, closely aligned with FHIR. It was believed that if a more generalized model were adopted in the future, the legacy data could be restructured to fit whichever model was chosen.

Steps in the development of the standard data extraction version 1:

1. Ad-hoc sql scripts.
    a. potentially different scripts for every project.
2. Standardised sql scripts for different purposes across multiple projects.
    a. Extractions for hospital admission from PIMs, medication orders from JAC, etc.
    b. Introduction of the concept of the research data view **(RDV).**
3. Standardising sql scripts across all output.
    a. Each extract had a standardised patient id, start and end time stamp, code and description for categorical data.
    b. Standardised definition of patient cohorts.
    c. Start and end dates of extraction.
4. Wrap sql scripts developed in previous step in python functions so that the different uses could be parameterised.
    a. Each function had to be called individually with their required parameters.
5. Input parameters and RDVs required via a CSV file.
    a. Single python call with the CSV file path as the only parameter.

Three main learning points

- Don't need a single source of data to create research data views.
    - Do require a single definition of a patient.
    - Don't listen to people saying it can't be done. What they mean is it is difficult to do.
- Learn from your customers.
    - Work out what's important to them and focus on that.
- Don't aim for perfection too early.
    - Have a try and see how it goes.
    - Learn from your mistakes.

GOSH Developed Tools, Deployed in Workspaces

The GOSH DRE Team developed R Shiny Apps, such as PICTURE and the Cardiac Dashboard, which are created once and shared across multiple research workspaces.

Cardiac M&M App

The “Cardiac M&M (morbidity and mortality) App” project is a dashboarding app which allows the Cardiothoracic surgery department of the hospital to see patient activity for defined metrics for their weekly multi-disciplinary meetings, for the purpose of managing patient care and improving departmental care standards. The data is extracted locally and pushed to the workspace, to which an RShiny app is then used to investigate the data through interactive tables and visualisations. Cardiothoracic clinicians have access to the workspace and can explore the data (including event history) as and when they please. The data is refreshed weekly which updates the outputs of the app.

A “Cardiac M&M de-ID App” has also been developed which reads in GOSH dummy cardiac data and displays a subset of the ‘real’ “Cardiac M&M App” features. The purpose of this app is for demonstrations to non-GOSH/ non-clinical audiences and for teaching purposes, as it does not contain any ‘real’ patient data - the data is simulated through GOSH’s dummy data processes .

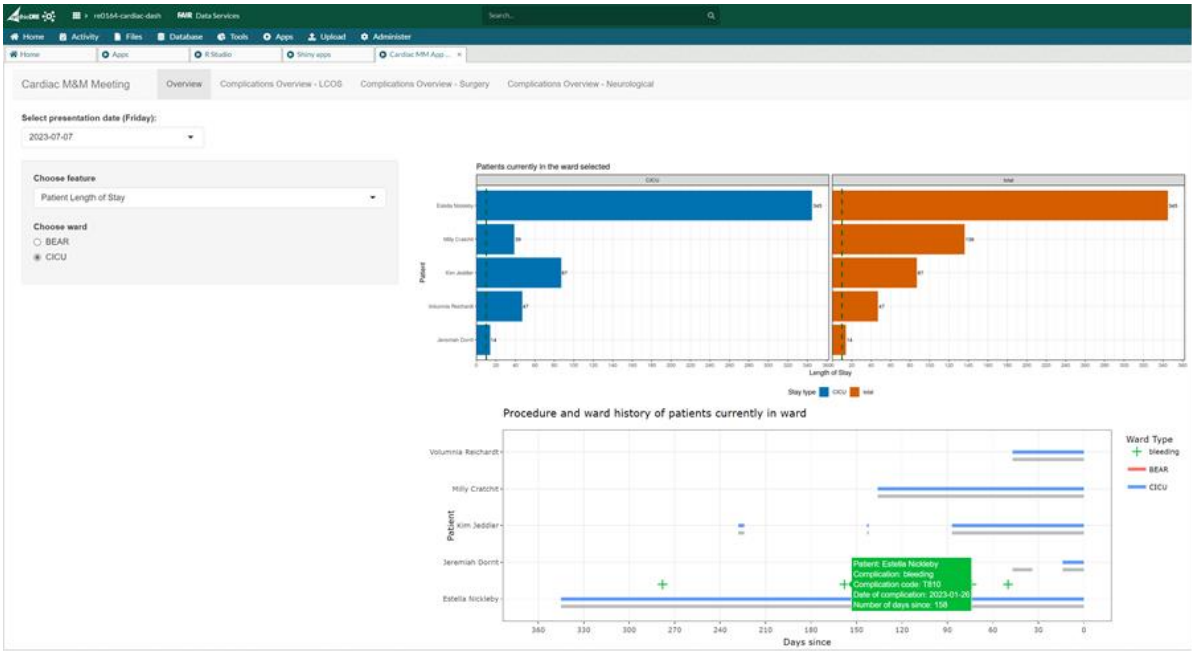


Figure 2 : Screenshot of "Cardiac M&M De-ID" RShiny App showing Patient Length of Stay in CICU and patient activity timeline.

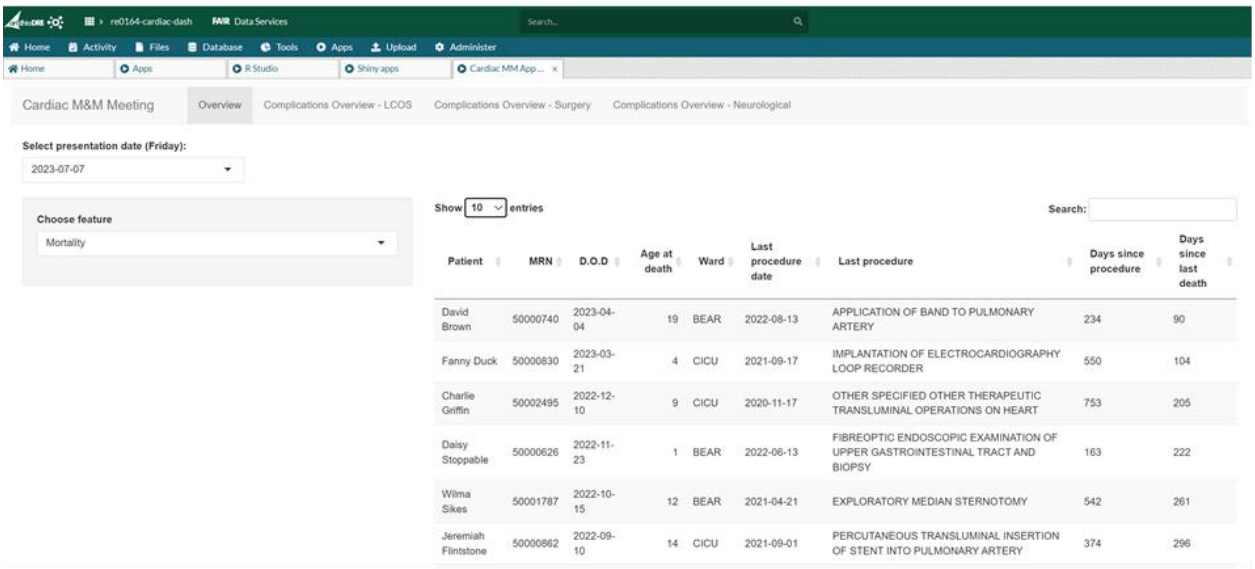
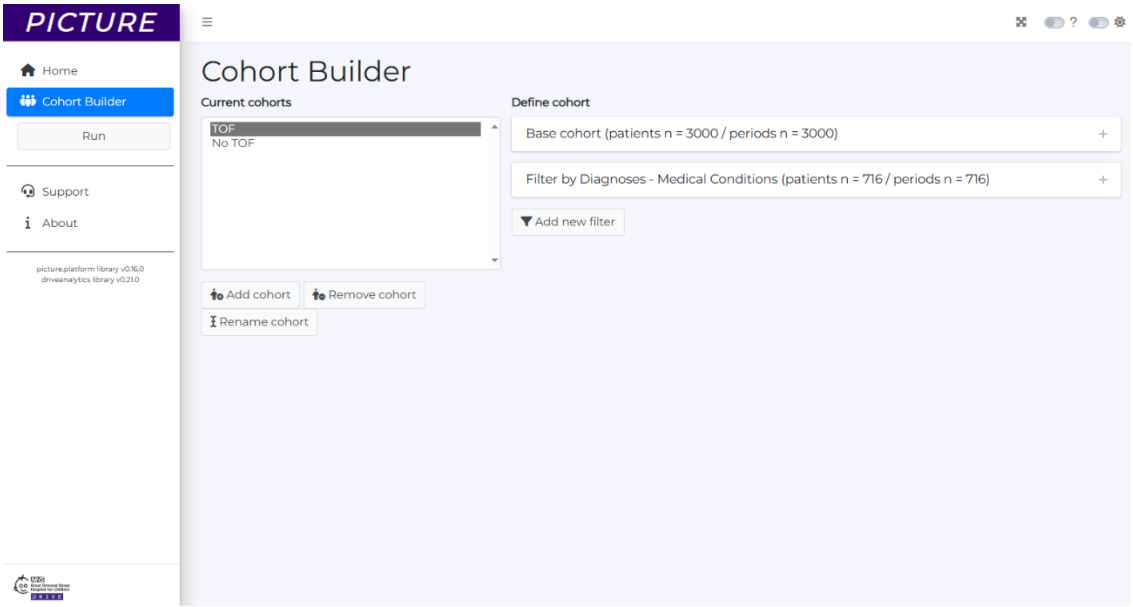
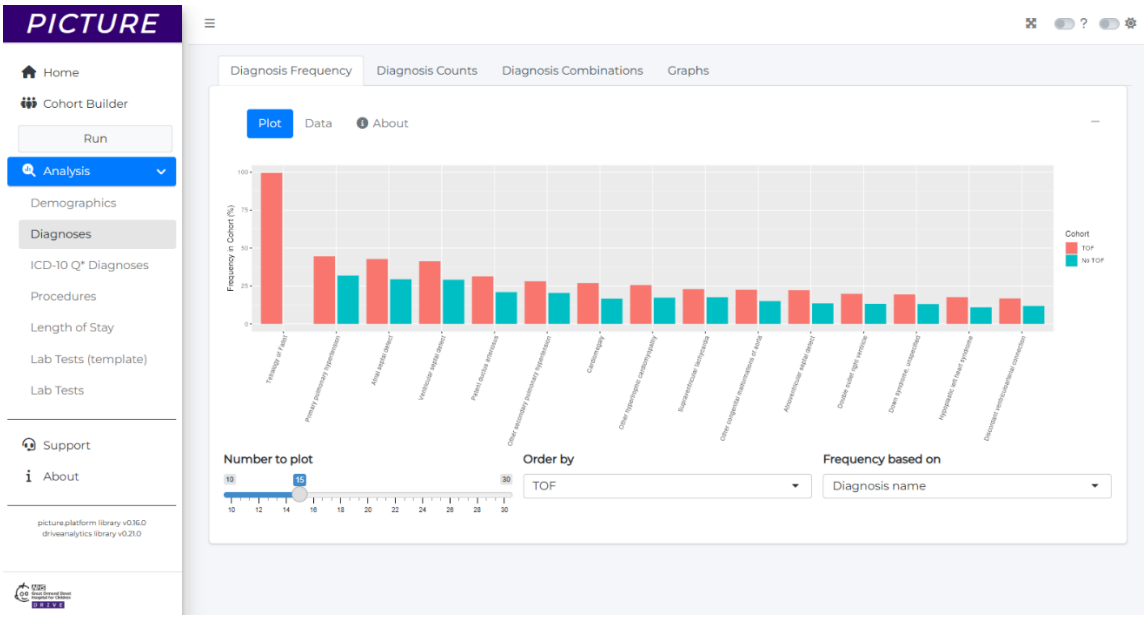


Figure 3: Screenshot of "Cardiac M&M De-ID" RShiny App showing event (mortality) history.

PICTURE App

PICTURE is a clinical informatics platform that takes in GOSH EHR data, offering a cohort builder, phenotype-based cohort segmentation, and standardized analyses. The platform generates either an interactive web app or a PDF report as output. It comprises a front-end R Shiny app and a separate package of analytic functions. PICTURE is installed in individual DRE workspaces and accessed through the Shiny app tab. To date, it has been used for pilot clinical use cases, tutorials and hands-on demonstrations. https://adc.bmj.com/content/108/suppl_1/A45.1





MLOPs in a Workspace

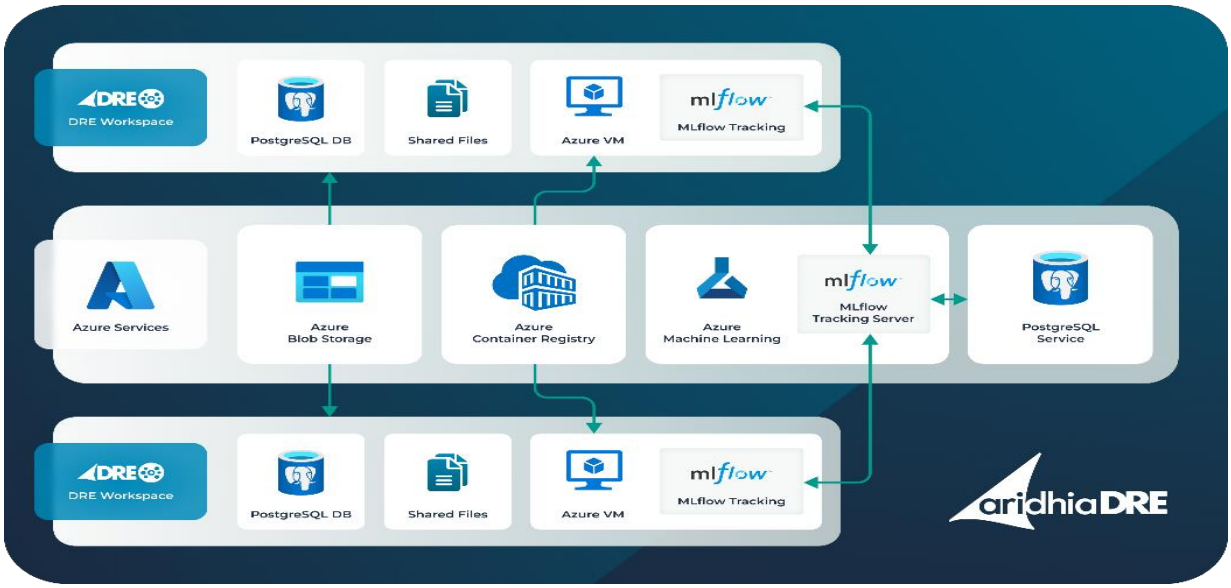
During 2024/25 Aridhia and GOSH have collaborated to bring MLOPs into the DRE. MLOps is essential for GOSH as they aim to operationalise AI effectively and achieve continuous improvement in their ML workflows. As part of this, the DRE supports model transparency and bias-awareness tools, enabling researchers to better understand model behaviour. Integration with explainability frameworks such as SHAP and LIME ensures that AI-driven insights remain interpretable and accountable.

So, what is MLOps? MLOps, short for Machine Learning Operations, is a set of practices, tools, and methodologies designed to streamline and standardise the deployment, monitoring, and management of machine learning models in production. It bridges the gap between data science and IT operations, enabling teams to collaborate effectively and ensure the reliability, scalability, and maintainability of machine learning systems.

We are working on two proofs of concepts (PoC)

Option A – Currently Available

This option is currently available within the Aridhia DRE and utilizes the VM/GPU in the project Workspace for processing. Each Workspace includes an ML Flow Tracking, which connects to a centralized MLflow Tracking server.

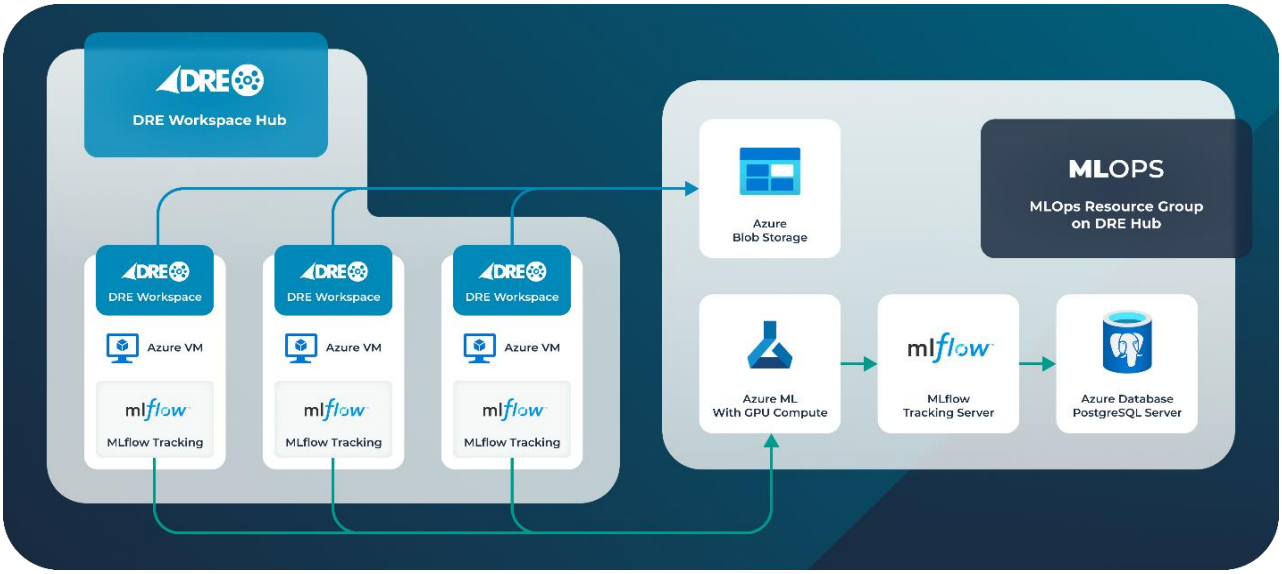


A dedicated resource group within the DRE Hub hosts the DRE Hub Azure Container Registry (ACR) for storing the completed containerised models. Azure Blob Storage is used for storing data, which can then be utilized to run the model. Workspaces can either share the same Blob Storage to enable data sharing or have separate Blob Storage instances mounted for isolated data access. Additionally, an Azure Database is used to manage MLflow admin logs.

Option B – In Progress

This PoC is currently in development and uses the Azure ML compute. Like Option A each Workspace includes an ML Flow Tracking, which connects to a centralised MLflow Tracking server.

A dedicated resource group within the DRE Hub provides access to the Azure ML compute. Azure ML also host the container registry for the completed models. Azure Blob Storage is used for storing data, which can then be utilized to run the model. As with Option A, Workspaces can either share the same Blob Storage to enable data sharing or have separate Blob Storage instances mounted for isolated data access. Additionally, an Azure Database is used to manage MLflow admin logs.



FHIR

In 2019, GOSH and Aridhia were selected as one of ten innovative healthcare data solutions in a UK-wide competition funded by HDR UK, tasked with answering a critical question: *How can we improve the sharing of clinical data to benefit patients both directly and indirectly?* This challenge formed the foundation of our Sprint Exemplar project, aimed at developing proof-of-concept technologies, methodologies, and research services.

The project focused on creating and demonstrating a functional exemplar of the complete health data research cycle at GOSH. This included metadata cataloguing, ontology mapping, de-identification, open-source reusable analytics, data auditability, and reporting—all implemented within the cloud-based DRE and applied to an example disease entity. A key outcome of this research cycle was the successful implementation of FHIR at DRIVE, data mapped to the FHIR standard and a new methodology for developing SMART on FHIR applications within a GOSH DRE Workspace.

Our approach tackled the problem from two main perspectives:

- **Research Enablement:** Securely and efficiently making routinely collected clinical data available to researchers while leveraging machine learning models.
- **Clinical and Patient Access:** Enabling clinicians and patients to access routinely collected clinical data, as well as other patient-generated data (e.g., from wearables and sensors), through specialised applications on smartphones and tablets. This was achieved using Fast Healthcare Interoperability Resources (FHIR) and SMART on FHIR (SoF) standards, with the application developed within the cloud-based DRE Workspace.

Since completing the Sprint Exemplar, GOSH now has two FHIR APIs integrated into their DRE. The first API contains synthetic data and is frequently used in hackathons. The second API holds real data and supports various projects, including the recent Genomics NLP project led by the GOSH DRE Team mentioned above.

The DRE Workspace has also been enhanced to support collaboration in developing SMART on FHIR apps within the environment, particularly when working with real-world data. Additions such as Git, integrated into built-in applications like Jupyter Notebooks and R Studio, and for development teams using a Virtual Machine, the inclusion of Gitea facilitates seamless version control and collaboration.

OMOP Data Transformation Project

In 2022, GOSH and Aridhia embarked on a new project to transform GOSH's Research Data Views (see Legacy Data & RDVs - How the GOSH DRE developed before EPIC. above) to the OMOP Common Data Model (CDM). For both organisations, this project represented their first engagement with transforming source data into the OMOP CDM, marking a milestone in their data standardization journey.

This initiative was partially funded by the European Health Data and Evidence Network (EHDEN), a program designed to promote interoperability and collaboration across the European health data ecosystem. In October 2021, Aridhia achieved certification as a Small and Medium Enterprise (SME) under the EHDEN program, further validating their expertise in health data transformation and analytics.

By standardizing selections of their source data to the OMOP CDM, GOSH has unlocked the potential to participate in federated research projects. This alignment with a widely accepted international data standard enables GOSH to benchmark their data against other hospitals while maintaining strict data governance by keeping sensitive information within their own infrastructure.

The project serves as a prime example of how adopting standardized data models like OMOP can drive both organizational and broader ecosystem benefits, such as enabling cross-institutional insights, improving research capabilities, and fostering collaboration across borders—all while safeguarding patient privacy. The OMOP curation environment was originally hosted on the DRE Hospital Landing Zone (HLZ) but has since been migrated to a DRE Hospital Workspace (HWS).

GOSH have now been involved in several Federation projects using the OMOP CMD including PHEMS which is detailed below.

A few of the key learnings:

Clinical Expert

Access to a clinical expert with in-depth knowledge of the data is essential for successfully transforming data to the OMOP Common Data Model (CDM). Their expertise ensures data quality

and minimises delays in the project. Mapping queries often required one to two weeks for resolution, slowing progress. Allocating a dedicated clinical data expert to the project for several days at the outset would significantly enhance efficiency and streamline the transformation process.

Infrastructure approach

At the outset of the OMOP project, the decision was made to host the OMOP Curation Environment within the Healthcare Landing Zone (HLZ), a secure, isolated section of the GOSH DRE. This approach was initially appropriate, as access was required only by a small, highly technical group from the DRE team.

Over time, a requirement emerged from GOSH ICT to implement multi-factor authentication (MFA) within this environment. While technically feasible, this solution was costly. In collaboration, GOSH and Aridhia developed an alternative approach, relocating the OMOP Curation Environment from the HLZ to a Hospital Workspace. This transition provided the DRE Team's Tenant Admins (TA) with full control to invite additional team members while leveraging the built-in MFA functionality. Additionally, this move resulted in cost savings.

Start Small

Early in the project, the GOSH and Aridhia teams recognised the importance of starting with a small scope. Mapping the RDVs to the OMOP Common Data Model (CDM) proved to be a time-consuming process, particularly as clinical expertise was not always readily available. Therefore, it was strategic to first complete the end-to-end transformation process for a limited number of RDVs before gradually expanding the approach.

Syndication with GOSH FAIR and HDR Health Data Gateway

Aridhia's FAIR technology suite is designed to support data custodians and researchers in implementing the principles of Findability, Accessibility, Interoperability, and Reusability. By providing tools for metadata standardisation, data cataloguing, and controlled access, Aridhia FAIR ensures that datasets are discoverable and usable for both humans and machines. **Syndication** is a critical component of this ecosystem, allowing datasets to be shared across platforms while maintaining robust governance controls.

The Health Data Research (HDR) UK Data Gateway is the central hub for discovering and accessing health data across the UK. Acting as a federated platform, it provides a single point of entry for researchers to locate datasets across a diverse network of contributing organisations.

GOSH has successfully integrated the FAIR component of their Data Research Environment (DRE) to serve as the authoritative source of truth for their dataset metadata. This strategic advancement

aligns GOSH with global best practices for data management and enables seamless metadata syndication to the HDR UK Data Gateway.

Federation Analysis and Learning – PHEMS Project

Aridhia and GOSH are both partners in the PHEMS project, a consortium of European paediatric hospitals exploring the use of data federation and synthesisation technologies to enable data sharing and research collaboration across national borders. The data standard used by this project is the OMOP CDM. The initial use cases the consortium is working towards are:

- Hospitals in the PHEMS network to share benchmarking data for clinical outcomes
- Hospitals in the PHEMS network pool their data to train machine learning (ML) models

To achieve the above Aridhia is further developing the Federated Node, an open-source software component for running federated analysis. The Federated Node can be integrated with the DRE, and GOSH are the first consortium partner to deploy this to their production infrastructure.

GOSH are leading the development of the federated analytics for the consortium, and with Aridhia's assistance will deploy these to their workspaces and Federated Node for testing. The current focus is the benchmarking use case, but development is underway on the PHEMS machine learning capability, with the intention of delivering an MVP in 2025. As part of this, the DRE supports model transparency and bias-awareness tools, enabling researchers to better understand model behaviour. Integration with explainability frameworks such as SHAP and LIME ensures that AI-driven insights remain interpretable and accountable.

Ongoing Collaborations

Aridhia and GOSH continue to collaborate on PoCs and projects such as MLOps in a Workspace, Imaging incorporating XNAT and data federation through the PHEMS project with a particular focus on AI projects.

Notes to Reader

About GOSH DRIVE:

GOSH Data Research, Innovation and Virtual Environments (DRIVE) Unit is the dedicated hub for innovation at Great Ormond Street Hospital for Children NHS Foundation Trust. The unit was established in 2018, following the implementation of the Electronic Patient Record system at the Trust. GOSH DRIVE's vision is to use data and technology in healthcare to better support staff and improve outcomes for children and young people.

Great Ormond Street Hospital Digital Research Innovation and Virtual Environment

About Aridhia Informatics:

Aridhia helps clinical and life sciences pioneers in Research Hospitals, Universities, Pharmaceutical Companies & Global Medical Networks who want to advance their medical research and improve outcomes more quickly.

Aridhia DRE | An Enterprise Scale Trusted Research Environment

What is the DRE?

The Aridhia Data Research Environment (DRE) is a cutting-edge digital platform designed to support secure, collaborative, and data-driven research in healthcare and life sciences. It provides a robust framework for managing, analysing, and sharing sensitive data while adhering to the highest standards of data governance, privacy, and security.

Core Purpose

The DRE facilitates innovation by enabling multidisciplinary research teams to work together seamlessly on complex datasets. It empowers researchers, clinicians, and data scientists to derive meaningful insights from data while ensuring compliance with regulatory frameworks such as GDPR and HIPAA.

Key Features

- **Secure Collaboration- Workspaces**:
  The DRE enables users to collaborate in a secure, Azure cloud-hosted workspace. Researchers from multiple institutions can access shared projects, ensuring that insights are generated collectively without compromising data security. The DRE supports a wide range of analytics workflows, from descriptive statistics and visualization to advanced machine learning and AI. Integrated tools and programming environments such as RStudio, Jupyter Notebooks and VMs where users can bring their own tools, make it highly versatile for both traditional and advanced data science tasks. The DRE has been praised for its ease of use.

- **FAIR Data Management**:
  Aligned with the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR), the platform offers advanced tools for metadata cataloguing, dataset search, and cohort building. The Aridhia FAIR Data Services make it easier to discover, understand, and manage datasets within a unified environment. FAIR Data Services encompass the entire process, from dataset discovery to request, approval, and delivery to a designated workspace.

- **Regulatory Compliance**:
  Designed with healthcare data requirements in mind, the DRE adheres to rigorous security and governance standards, enabling the safe use of sensitive data for research purposes. It ensures compliance with legal requirements, allowing researchers to focus on generating insights. To ensure responsible AI adoption, we are investigating the platform integrating AI governance frameworks, offering researchers bias-detection tools, explainability features, and auditability for AI-driven research projects. These features will ensure compliance with evolving regulatory standards while supporting the ethical and transparent use of AI in healthcare and research.

- **Scalability and Flexibility**:
  Built on the Azure cloud infrastructure, the DRE is scalable to meet the needs of diverse research projects, from single-centre studies to multi-institutional collaborations. Its modular design allows it to adapt to different research needs, offering specialized workspaces for hospitals, academic institutions, and research consortia.