foundry120

PLATFORM DOCUMENTATION

Platform Reference Guide

Architecture, governance, and operational overview

AUTHOR	DATE
Dr Shaun Chuah	1 May 2026

VERSION

v1.0

foundry120

Table of Contents

Document status and review note
This guide is intended to support adoption and governance review. It describes the current Foundry120 platform architecture and operating model but does not replace local sponsor, REC, DPIA, information governance, data protection, or institutional review where required. Adopting teams remain responsible for confirming that their proposed use of Foundry120 is compatible with applicable study approvals, consent language, sponsor expectations, institutional policies, and local governance requirements.

1. Overview

Foundry120 is a governed translational research platform designed to help biomedical research teams organise samples, centralise research data, and accelerate discovery with AI-assisted analysis. It began as a sample tracking system for the MUSIC IBD study but has expanded into a broader research operating system for multi-study, multi-site translational science.

Foundry120 brings together three connected layers:

- **G-Trac** is the sample and study tracking layer. It supports participant-linked sample records, QR-code workflows, freezer and box management, sample movement history, experiments, audit trails, and role-based access.
- **Orca** is the data orchestration layer. It ingests research data from external systems such as REDCap, transforms and validates datasets, documents lineage, and publishes cleaned outputs into the governed platform. Data can be onboarded through automated integrations or structured manual uploads depending on the needs of the adopting team.
- **Helix** is the AI-assisted analysis layer. It allows researchers to ask plain-English questions over approved datasets, documents, and sample records. Helix can search, summarise, chart, join uploaded files to governed data, and support controlled analysis workflows.

Why It Exists

Translational research teams often work across fragmented systems: spreadsheets, freezer logs, REDCap exports, local files, protocols, and ad hoc analysis scripts. This makes it difficult to know what samples exist, what data are linked to them, who can access what, and how previous work can be reused.

Foundry120 aims to solve this by creating one governed environment where samples, clinical data, omics, documents, and analytical workflows become traceable, queryable, and reusable across studies, sites, and disease areas.

What It Enables

Foundry120 helps research teams:

- Track samples end-to-end with clear audit history.
- Link biospecimens to clinical, omic, document, and derived research data.
- Manage access through role-based governance.
- Transform raw study exports into reusable datasets.
- Search and analyse approved research data using natural language.
- Reuse prior data and specimens more effectively across future studies.

Long-Term Vision

The long-term goal is for Foundry120 to provide repeatable digital infrastructure for translational research: turning fragmented samples, datasets, documents and analytical outputs into governed, reusable research assets.

2. Access Control

Foundry120 uses scoped role-based access control. Every user's access is described by two things: a role (what they can do) and a scope (where that applies). A user with study operator access on MUSIC cannot see MARVEL data unless they are explicitly assigned there. This is by design.

The two primary scopes

Most day-to-day access is managed through two scope types:

- Study — a research protocol (e.g. MUSIC, MARVEL). Governs samples, study identifiers, and study-linked datasets.
- Lab group — a PI/research team (e.g. Bain lab). Governs freezer boxes, experiments, and lab-governed datasets.

Organization and platform scopes exist for institutional user administration and shared reference data respectively.

Study dataset access is a separate gate

Operational study access (viewer/operator/admin) does not automatically grant access to study datasets. Dataset access requires an explicit **Study Data Reader** role assignment on that study. This is intentional: the team member who registers samples does not automatically receive access to governed analytical data. The two are provisioned independently. Lab operators and leads receive lab dataset access automatically as part of their operational role, because lab data and wet-lab work are considered part of the same workflow. This gate also governs what data Helix can access on behalf of a user — Helix queries and tool calls operate within the same scoped permissions.

Role ladders

Study access follows three levels:

Role	What they can do
Study Viewer	Read samples and study identifiers
Study Operator	Create and update samples and study identifiers
Study Admin	Everything above, plus manage study access and govern dataset gates for that study

Lab group access also follows three levels:

Role	What they can do
Lab Viewer	Read boxes and experiments
Lab Operator	Create and manage boxes, experiments, and lab-governed datasets
Lab Lead	Everything above, plus manage lab access and govern lab dataset workflows

Who can manage access

Admins and leads can create and remove role assignments within their own scope:
Study Admin — manages study access (all study roles, including data reader) for their study
Lab Lead — manages lab group access for their lab. No one can grant access beyond their own scope.
A study admin for MUSIC cannot grant MARVEL access.

Composability

Users can hold multiple roles simultaneously. A common example: a researcher who is a study operator on one study, a lab operator in their team's lab, and holds data reader access on an approved study dataset. Foundry120 combines those permissions at runtime — the user sees all the areas they need.

Least-privilege default

If no relevant role binding exists for a resource, access is denied. The platform does not grant partial or implied access. If something is missing from a user's view, the fix is to assign the correct scoped role.

3. Data and Security

Foundry120 is designed to hold pseudonymised research data only. This section describes how data are organised, where infrastructure is hosted, and what security and resilience controls are in place.

Data model

Data are organised around governed research objects:

- studies and lab groups
- participants and study identifiers
- samples and storage locations
- experiments
- datasets, derived outputs, and computational workflows
- documents and protocols

This lets research data be linked and queried without relying on informal spreadsheets, duplicated exports, or uncontrolled local files.

Datasets

Datasets uploaded into Foundry120 are treated as governed research assets. Each dataset should carry sufficient metadata to describe what it contains, which study, samples, or participants it relates to, who may access it and where it came from. The intention is reusability and traceability, not simply file storage.

Data exports

Export capability is governed by scoped access controls: users can only export data they are authorised to access. Helix can generate derived outputs such as CSV files, joined datasets and PNG charts for downstream analysis. Because Foundry120 is designed for pseudonymised data, exports generally carry a lower risk profile than exports from identifiable clinical systems, but they still require study-specific governance, PI approval and institutional controls.

Infrastructure and storage

Foundry120 is deployed within the University of Glasgow's institutionally managed Microsoft Azure infrastructure. Platform services, database infrastructure, dataset storage, and audit records are hosted within this governed cloud environment.

Core platform data storage remains within the UK South Azure region. Production Helix inference is configured through Microsoft Azure AI Foundry with processing restricted to EU Azure regions. This means core datasets remain hosted in UK South, while selected data submitted for Helix inference may be processed within Microsoft-managed EU Azure infrastructure.

Security controls

Foundry120 uses layered security controls:

- role-based access control with least-privilege defaults
- study and lab-group scoping
- separate dataset access gates
- audit logging
- encrypted transport and storage

- controlled deployment and restricted administrative access

Auditability

The platform maintains records of data access, sample movement, dataset registration, and workflow activity. This supports accountability for data use, traceability of derived outputs, and governance review where required.

Backup and resilience

Foundry120 uses Azure-native resilience controls, including managed backup mechanisms, geographically redundant storage, blob versioning, and soft-delete protections.

The platform database contains critical operational metadata, including access control records, study/sample relationships, dataset provenance, audit records and workflow state. In addition to Azure-native backup mechanisms, database backups are maintained in geographically and administratively separated environments to support recovery from major infrastructure-level incidents.

Foundry120 is primarily a governed analytical and computational research environment for pseudonymised research datasets. Long-term archival arrangements for primary raw research data, including raw data from laboratories, sequencing, imaging, or external repositories, remain the responsibility of the adopting team and relevant institution.

4. Governance

Adoption involves shared responsibility between the Foundry120 platform team, the adopting study team, and the relevant sponsor or institution. This section covers the governance position for computational analysis, and how responsibility is divided.

Helix and computational assistance

Helix is a computational assistance layer within Foundry120. It supports querying approved datasets, summarising content, generating charts and tables, analysis planning, code generation, and controlled analytical workflows.

Helix does not replace investigator judgement and is not intended to make autonomous clinical or scientific decisions. Outputs are assistive and should be reviewed by appropriately qualified researchers.

Helix operates within the same scoped access controls as the rest of the platform. A user's Helix session can only retrieve or analyse data that the user is already authorised to access. User queries to Helix are logged. Tool calls that access governed data, such as dataset lookups, sample searches or analytical operations, are logged separately to provide an auditable record of data access.

Pseudonymised data

Foundry120 is intended for pseudonymised research data only. Direct identifiers must not be uploaded. This is a platform-level requirement, not a recommendation. This distinguishes Foundry120 from systems that process identifiable clinical records, direct care data, or unfiltered clinical documentation, and substantially reduces the governance risk profile.

When additional review may be needed

Additional governance review may be appropriate where a proposed use involves:

- identifiable participant information
- cross-study analysis beyond existing approvals
- new secondary uses not described in the original protocol
- external collaborator or commercial partner access
- participant-level automated interpretation
- outputs intended to inform direct clinical care

The need for additional review is driven by the proposed use case, not simply by the presence of AI-assisted tooling.

Shared responsibilities

Party	Responsibilities
Foundry120 platform team	Platform infrastructure and security; access control mechanisms; audit logging; Helix inference configuration through EU-region Azure services; dataset onboarding support; platform documentation and technical support.
Adopting study team	Confirming that proposed use is compatible with protocol, consent, REC approval, sponsor expectations and local governance; ensuring uploaded

	data are pseudonymised; approving users; reviewing outputs; managing exports and long-term archival arrangements.
Sponsor and institution	Study sponsorship oversight; formal data governance and data protection compliance; institutional cloud and infrastructure policies; approval of external collaborations; retention and archival policies.

Current boundaries

Foundry120 is not currently intended for direct clinical care, storage of identifiable participant records, autonomous clinical decision-making, or unrestricted external data sharing. Any proposed use outside the standard pseudonymised translational research workflow should undergo additional local governance review.

5. Use Cases

The following table is intended to help teams review proposed uses. It is not a substitute for local sponsor, REC, DPIA, or institutional review where required.

Use case	Typical review position	Notes
Searching for available pseudonymised samples	Often compatible with existing approvals	e.g. identifying available serum, stool, tissue, or PBMC samples.
Querying pseudonymised study metadata and cohort characteristics	Often compatible with existing approvals	Assumes the query remains within the approved study purpose.
Uploading clean pseudonymised datasets for analysis	Often compatible with existing approvals	Appropriate where computational analysis is within the study scope.
Generating code, charts, tables, or exploratory summaries	Often compatible with existing approvals	Outputs remain investigator-reviewed.
Registering derived outputs back into the platform	Often compatible with existing approvals	Supports provenance and reuse.
Linking samples to approved clinical, omics, or derived data	Often compatible with existing approvals	Depends on consent and study scope.
Generating statistical models, ML classifiers, or derived analytical outputs from governed datasets	Often compatible with existing approvals	Outputs remain investigator-reviewed. Any modelling should use only approved pseudonymised datasets.
Exporting derived spreadsheets or charts from Helix for external analysis	Often compatible with existing approvals	Governed by scoped access controls. Applies to pseudonymised derived outputs only.
Cross-study analysis across multiple cohorts	May require additional review	Depends on consent, protocol scope, sponsor position, and data-sharing arrangements.
External academic collaborator access	May require additional review	Usually requires PI approval and appropriate collaboration arrangements.
Commercial partner access	Requires specific governance review	Should not be assumed to be covered by standard academic research approvals.
Uploading directly identifiable participant data	Out of scope	Foundry120 is designed for pseudonymised research data only. Direct identifiers must not be uploaded.
Direct clinical decision-making or autonomous clinical care	Out of scope	Foundry120 is a translational research platform, not a clinical care system.

Adoption patterns

A lower-risk adoption pattern typically involves pseudonymised datasets only, existing approval covering translational or computational analysis, access limited to the approved study or lab team, no identifiable free text, no direct clinical decision-making, and investigator review of all outputs.

A higher-scrutiny pattern includes one or more of: broader secondary use, cross-study pooling beyond existing approvals, external data sharing or commercial collaboration, automated participant-level interpretation, or uncertainty about consent scope.

6. Getting Started

This section describes the typical onboarding pathway and answers common questions teams ask before adoption.

Onboarding pathway

1. **Initial discussion.** The adopting team discusses the proposed use case with the Foundry120 team, including study context, dataset types, expected users, and governance constraints.
2. **Local governance review.** The adopting team reviews whether the proposed use is compatible with the study protocol, consent language, REC approval, sponsor requirements, and institutional policies. The Foundry120 team can provide documentation to support this review.
3. **Workspace setup.** Once governance alignment is confirmed, the Foundry120 team configures the relevant workspace, including study or lab group structure, user roles, dataset categories, access gates, and metadata fields. This can proceed in parallel with governance review where the adopting team is comfortable.
4. **User provisioning.** Users are added with scoped roles. Access is limited to individuals with a legitimate role in the relevant study, lab group, or collaboration.
5. **Dataset onboarding.** Datasets can be brought into Foundry120 through automated Orca ingestion pipelines or structured manual upload. Automated pipelines may be appropriate for recurring feeds from systems such as REDCap, while manual upload may be sufficient for one-off or less frequently updated datasets. In both cases, datasets are assessed for structure, pseudonymisation, provenance and compatibility before being registered as governed platform assets.
6. **Operational use.** Teams use Foundry120 for sample tracking, dataset discovery, cohort querying, computational analysis, and Helix-assisted workflows.

Teams should reassess governance requirements when new datasets or collaborators are added, the study purpose changes, cross-study analysis is proposed, or commercial involvement is introduced.

Frequently asked questions

Does Foundry120 store identifiable participant data?
No. Foundry120 is designed for pseudonymised research data only. Direct identifiers must not be uploaded and are outside the supported operating model.

Who approves user access?
User access is approved within the relevant study or lab group governance structure. Study admins and lab leads can manage access within their own scope but cannot grant access outside that scope.

Where is data stored and where does Helix processing occur?
Datasets are stored within institutionally managed Microsoft Azure infrastructure. Production Helix inference is configured through Microsoft Azure AI services, with processing restricted to EU Azure regions.

Are research data used to train AI models?
No. Research data processed through Helix are not used to train external foundation models.

Does using Foundry120 require REC amendment or reconsent?
Not automatically. Many workflows involving computational analysis of pseudonymised research data may already fit within existing approvals. Additional review may be needed for new secondary uses, cross-study pooling, identifiable data, commercial access, or clinical decision-support applications. The adopting team, sponsor, and institution make this determination.

Is Helix making autonomous decisions?
No. Helix provides computational assistance. Outputs remain subject to investigator review and do not replace scientific, statistical, clinical, or governance oversight.

Is Foundry120 a clinical system?
No. Foundry120 is a translational research platform. It is not intended for direct clinical care or autonomous clinical decision-making.

Can data be exported from the platform?
Data export should follow study-specific governance, PI approval, and institutional requirements. Helix can generate exports of query results, joined datasets, and charts. These are governed by the same scoped access controls as platform data.

Is Foundry120 the long-term archive for all raw research data?
Not by default. Foundry120 is primarily a governed analytical and computational research environment. Primary archival responsibility for raw data from laboratories, sequencing facilities, imaging platforms, or external repositories remains with the adopting team and institution.