GUARDIANS CCIA ZERO Data Commons – API Requirement Specifications Report

Contents

1. Purpose

This document outlines the scope of the initial API requirements for CCIA's ZERO Data Commons project delivered under the GUARDIANS initiative. It summarises the list of requirements for data fetching and evaluates whether existing infrastructure can already support them.

Note that this is a high-level, baseline artefact reflecting the current state of development as of November 2025. The requirements presented here are derived from preliminary user stories and stakeholder input, and are expected to evolve as platform development progresses and further feedback is gathered. Some requirements are currently satisfied by existing infrastructure (e.g., FHIR API), while others will require further investigation and custom development (see section 2 below). Where feasibility is marked as "partially achievable" or "not achievable", these items are flagged for future technical review and prioritisation.

This document is not exhaustive, but establishes a foundation for ongoing API design, testing, and documentation. More detailed specifications, including technical implementation notes and validation criteria, will be delivered in subsequent milestones. We welcome feedback and will continue to update this artefact as the project advances.

2. API Requirements

We defined several requirements for programmatic access to data within the data commons, derived from the preliminary user stories documented early 2025. The table below outlines the specific requirement definition and whether it is satisfied with our current infrastructure.

Req#	Requirement	Feasibility	Rationale	Linked User Story	Next Steps
API#01	Fetch all patients	Achievable with FHIR API	Foundational cohort discovery enabling counts, sampling frames, and baseline analytics for ZERO Data Commons.	US01 - Get all withdrawn patients	Confirm Patient search/scopes on FHIR; implement pagination & server-side filtering; validate de-identification for research contexts.

Req#	Requirement	Feasibility	Rationale	Linked User Story	Next Steps
API#02	Fetch all patients and their related clinical metadata	Achievable with FHIR API	Combines clinical context with patient lists to enable richer stratification and longitudinal insights.	US04 - Get methylation class and score for all tumour patients	Map ZERO fields to FHIR resources (Condition, Observation, Encounter); define “related metadata” scope; add include/expand parameters.
API#03	Fetch all reportable molecular variants for a given patient	Achievable with FHIR API	Supports clinical decision-making, tumour board prep, and audit trails of reportable findings.	US07 - Search full analysis (full suite of data for a given patient) set by any id (tumour, normal, manifest, rnaseq, etc)	Ensure linkage Patient to Specimen to Molecular findings; codify “reportable” classification; create patient-scoped endpoints/queries.
API#04	Fetch all patients with at least one sample with WGS data	Achievable with FHIR API	Enables modality-driven cohort building and WGS utilisation metrics.	US11 - Get number of patients that have done WGS	Define modality flags (WGS/WTS/methylation) in data model; index sample modalities; add filters for “≥1 WGS”.
API#05	Fetch all samples that showcase at least one specified reportable molecular variant	Achievable with FHIR API	Critical for frequency analyses, variant-of-interest triage, and cross-cohort comparisons.	US05 - Get all samples with a specific variant	Implement variant-indexed search; define “specified reportable” filter schema; validate performance on large variant sets.
API#06	Fetch all PDX models for a given cancer category	Partially achievable with FHIR API – will require further investigation	Connects clinical cases to pre-clinical models for translational research and pipeline testing.	US13 - Get all samples with PDX or HTS data	Confirm PDX representation in schema; design bridging model (Databricks or custom endpoint); harmonise cancer-category ontology.

Req#	Requirement	Feasibility	Rationale	Linked User Story	Next Steps
API#07	Fetch all samples that showcase a COMBINATION of specified reportable molecular variants	Not achievable with FHIR API – will require investigation into Databricks and custom API endpoints	Enables co-occurrence discovery, resistance mechanism hypotheses, and subtyping.	US08 - Get all samples (given a specific diagnosis eg DMG) with a specific combination of mutations	Design composite-variant query service; prototype Databricks SQL endpoints; define AND/OR logic & performance constraints; add integration tests.
API#08	Fetch all molecular variants, regardless of pathogenicity	Not achievable with FHIR API – will require investigation into Databricks and custom API endpoints	Supports exploratory research (e.g., VOUS), trend detection, and future reclassification.	US17 - Get all samples with a specific class importance (VOUS3.8, etc)	Expose complete variant catalogue endpoint; include classification attributes; batch export; provenance tracking.

3. Next steps

The table above represents a baseline for the API requirements in 2026; as such, it is expected to change in the coming months as development of the Insights platform progresses. The table below outlines the next steps and planned stages related to the API requirements. Web interface development will continue in parallel, with each stage informed by user feedback and testing.

Stage	Target date	Description
Defined API requirements	November 2025	As described in this document
API development	May 2026	Secure API development: Implementation and testing of APIs for basic data access and sharing, compliant with FAIR principles. API documentation and release: Open-source publication of fully documented APIs. Semi-automated DAC system: Deployment of data access request system, validated against user stories and access governance needs.
Production release	November 2026	Finalised web interface, data model, APIs and DAC system delivered as a fully integrated platform.