

HCMI Software Architecture

Version History

Version	Implemented By	Revision Date	Approved By	Approval Date	Description of Change
1.0	Rahul Verma / Francois Gerthoffert	April 27th, 2018	Vincent Ferretti	-	Initial Version
1.1	Rahul Verma	May 9th, 2018	Vincent Ferretti	-	Updated installation instructions
1.2	Rakesh Mlstry	July 14th, 2023	Scott Cain	-	Updated to reflect current architecture

Table of Contents

Introduction	4
Software Architecture	5
Overview	5
Breakdown	5

Introduction

The Human Cancer Model Initiative (HCMI) Searchable Catalog provides a web based interface that allows the biomedical research community to search for and inspect next-generation tumor models that have been made available to the community by the Office of Cancer Genomics (OCG).

The HCMI Software Architecture document is intended to provide stakeholders and technical teams with an overview of all components implemented and deployed for the project.

Software Architecture

Overview

The system is composed of two major software components:

- A public-facing Catalog UI providing users with the ability to search and view models
- An access-controlled content management system (CMS) to manage data available in the Catalog

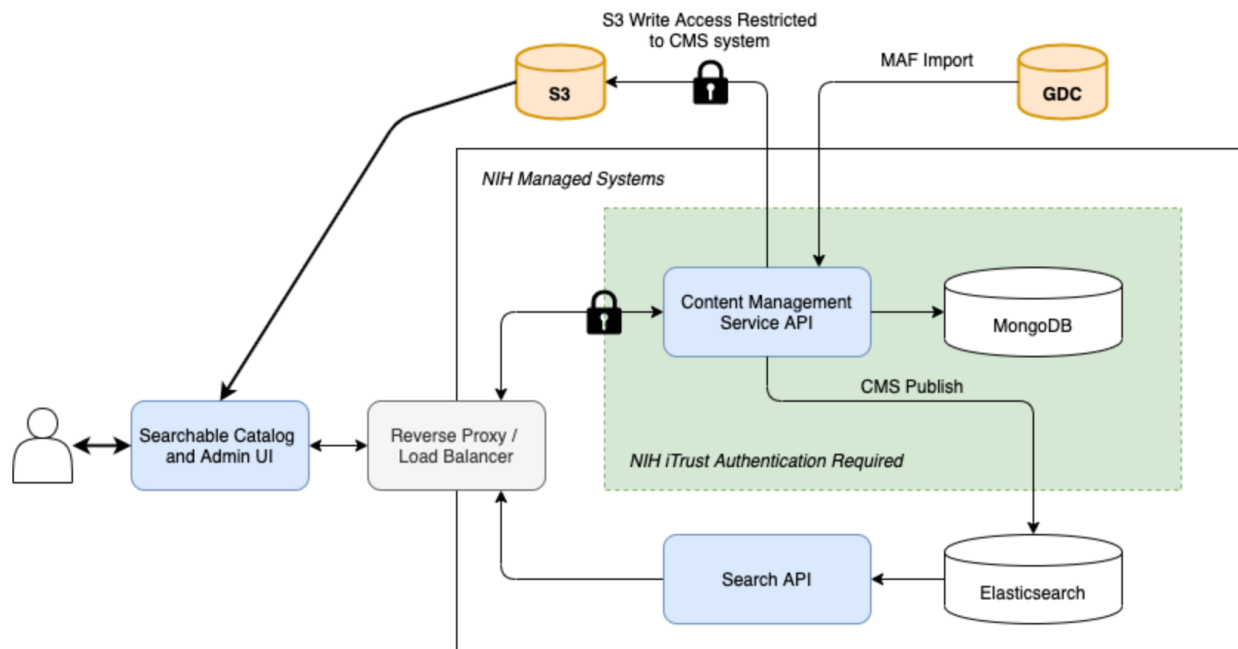


Figure 1: HCMI System Architecture Diagram - The latest microservice architecture is decomposed naturally along the problem domain. Services in blue handle their respective areas of the problem domain.

Breakdown

The HCMI Searchable Catalog is composed of four loosely-coupled service components that each functions as an expert in its domain, as described below:

- **Content Management Service (CMS):** A backend service that is access controlled via NIH iTrust and allows adding, modifying and deleting of the models, registering of Content Managers and importing variant data from the GDC. This component is written in JavaScript, using Express.js.
- **Publishing Service:** A backend service that publishes new or updated cancer models to Elasticsearch.
- **Search API:** A backend service optimized specifically for quickly searching the Catalog and responding to web queries. This service is used for all of the public,

non-administration pages. This component is written in JavaScript, using Express.js and using the Overture component called Arranger (<https://github.com/overture-stack/arranger>).

- **Searchable Catalog and Administration UI:** The web interface that users interact with providing seamless access to all backend services. This is written in Javascript using ReactJS.

The above outlined components are supported by the following external data services:

- **MongoDB** - Used to store model data. This data can be added or edited by the CMS.
- **Elasticsearch** - Provides searchable data for the catalog. Data that the CMS publishes using the Administration tools are stored in Elasticsearch for the Search API to quickly and easily access.
- **S3 Bucket:** An AWS S3 object storage bucket used to store images. HCMI backend services use this as a source to serve image content to the Searchable Catalog.
- **Genomic Data Commons (GDC):** A unified data repository that enables data sharing across cancer genomic studies. The HCMI application uses the GDC as a source of variant data for the stored models.