

# Alignment Engine

We believe that it is important to keep the databases mostly unaware in which format the data was originally stored. A reference to this format will only be stored for specific purposes involving file transfers.

Data model for variants and alignments have been designed and implemented in Java. They explicitly specify the most commonly used fields, and at the same time provide mechanisms for preserving all the information of a certain format. For instance, the fields specified for a variant would be (among others) chromosome, position, reference and alternatives; if a VCF file is being stored, then columns such as INFO are also saved in a key-value data structure.

OpenCGA imports different data models from [OpenCB Biodata](#) and [GA4GH](#) such as Variant and Alignment data models; while others such as [Catalog Data Models](#) have been developed in OpenCGA itself. In next sections you will find

## Table of Contents:

- [Catalog](#)
- [Storage](#)
  - [Variant](#)
  - [Alignment](#)

## Catalog

*Catalog* models all the information about users, projects, studies, files, jobs, samples and clinical data among others. This has been developed internally in OpenCGA Catalog component, you can find a more detailed information at [Catalog > Catalog Data Models](#).

## Storage

### Variant

OpenCGA Variant data model has been developed in [OpenCB Biodata](#) and is used in different OpenCB projects such as [CellBase](#) or [Oskar](#).

### Alignment

OpenCGA takes Alignment data model specification from [GA4GH](#) and the implementation from [OpenCB GA4GH](#). See a full description at [Alignment Data Model](#).