

Article

A Hybrid PACS/FHIR Model for Clinical Genomic Data Management of Cardiomyopathies and Channelopathies: Standardized VCF Archiving for Hereditary Heart Disease

Paolo Marcheschi ^{1,*}, Stefano Dalmiani ¹, Maria Pisani ¹, Martina Modena ^{2,3}, Gunter Zeilinger ⁴, Vrinda Nayak ⁴, Divon Mordechai Lan ⁵, Simona Vittorini ³, Nicoletta Botto ³ and Marcel Swennenhuis ⁴

¹ Reparto BITE, G. Monasterio Foundation, CNR-Regione Toscana, via G. Moruzzi 1, 56124 Pisa, Italy; stefano.dalmiani@ftgm.it (S.D.); mpisani@ftgm.it (M.P.)

² Interdisciplinary Center for Health Sciences, Scuola Superiore Sant'Anna, 56127 Pisa, Italy; martina.modena@santannapisa.it (M.M.)

³ Molecular Cardiology Laboratory, G. Monasterio Foundation, Ospedale del Cuore, CNR-Regione Toscana, via Aurelia Sud, 54100 Massa, Italy; simona.vittorini@ftgm.it (S.V.); nicoletta.botto@ftgm.it (N.B.)

⁴ J4care GmbH Enzersdorfer Straße 7, 2340 Mödling, Austria; gunter.zeilinger@j4care.com (G.Z.); vrinda.nayak@j4care.com (V.N.); marcel.swennenhuis@j4care.com (M.S.)

⁵ Genozip Limited, Hong Kong, China; divon@genozip.com (D.M.L.)

* Corresponding author. E-mail: paolo.marcheschi@ftgm.it (P.M.)

Received: 18 March 2026; Revised: 10 July 2026; Accepted: 13 August 2026; Available online: 30 September 2026

ABSTRACT: The precise diagnosis and lifelong management of hereditary heart diseases, particularly cardiomyopathies and channelopathies, are increasingly dependent on the rapid analysis of massive genomic data generated by Next-Generation Sequencing (NGS). This reliance creates a critical challenge: securely and efficiently integrating large Variant Call Format (VCF) files into existing hospital information systems. This technical hurdle must be overcome to realize the promise of personalized cardiovascular medicine. The ability to sequence the human genome has ushered in a new era of personalized medicine, particularly within the field of precision cardiology. By gaining a thorough understanding of each patient's genetic profile, physicians can now move beyond generalized care to tailor treatment plans to the individual. In clinical follow-up, genomic profiling has become essential to: (1) Diagnose genetic disorders, specifically Inherited Arrhythmias and Cardiomyopathies. (2) Perform family cascade screening to identify at-risk relatives. (3) Contribute to risk stratification for Sudden Cardiac Death (SCD) in the genes for which gene-specific risk models have been validated. (4) Guide life-saving interventions, such as determining the precise timing for Implantable Cardioverter-Defibrillator (ICD) placement. The ability to quickly and easily access all pertinent patient data is the main benefit of an Electronic Health Record (EHR). In this work, we will show how electronic health record systems (EHRS) can access and use genomic information in clinical care, using compression to enable fast, standardized storage and retrieval of data with tools available in a standard hospital setting, and leveraging the corresponding PACS functionality. By storing genomics data

in a central repository accessible to EHR systems, healthcare providers can access the data when they need it. This can help to shorten the time spent diagnosing and treating patients. This approach was experimentally implemented and evaluated using the clinical genomic dataset from the FTGM Molecular Cardiology Laboratory, which routinely performs NGS for patients with hereditary heart disease. The J4CARE VNA PACS, based on the Open Source DICOM archive DCM4CHEE Archive 5.x, was selected as the repository for this experimental platform.

Keywords: Electronic Health Records (EHR); Picture Archiving and Communication Systems (PACS); DNA sequencing; Personalized medicine; Genomic information; Clinical care; Fast storage and retrieval of data; Standardized way of storing data; Compression; Central repository for genomics data; Efficiency of healthcare delivery

1. Introduction

Genomic medicine is on the cusp of transforming healthcare, but its promise hinges on the seamless integration and availability of DNA data into Electronic Health Record Systems (EHRS). While recent advances have made genome sequencing more accessible, EHR systems largely remain unequipped to handle the scale and complexity of genetic information [1,2].

This work is specifically grounded in the clinical workflow of the FTGM Molecular Cardiology Laboratory, which performs high-throughput NGS for approximately 600–700 patients annually with suspected hereditary heart diseases, specifically cardiomyopathies and channelopathies. The resulting large volume of VCF data generated by this high-cadence clinical service provided the critical foundation and imperative for developing a standardized, scalable genomic archiving architecture.

Representing or linking EHR DNA data in a meaningful way requires addressing pragmatic (clinical), ethical, and technical challenges.

On the pragmatic front, deciding what data to include and how to structure it poses issues. Basic patient-centered approaches that store raw sequence data may overwhelm systems, whereas pointer-based methods that link out to external databases raise interoperability concerns [3]. Standardizing genomic data types and developing new data models will be key.

Ethically, properly safeguarding sensitive DNA information and ensuring confidentiality are paramount [2,4].

Ideally, an evolved EHR system with the functionality to manage genomic data requires advanced data integration, representation, and decision-support tools that can link genotypes to phenotypes and provide clinical guidance, using EHRS capabilities for receiving, storing, and presenting complex DNA data[1].

The clinical adoption of massively parallel sequencing rests on three decades of technological development, from the introduction of automated fluorescence-based DNA sequencing to the current generation of high-throughput platforms now routinely deployed in diagnostic laboratories [5–7]. As sequencing has become accessible, the limiting factor has shifted from data generation to data integration, a problem addressed by professional bodies on both the informatics and the clinical genetics side [8,9], and clinical experience has shown that structured genomic and phenotypic data held within the EHR can materially improve diagnostic yield in exome-based analyses[10]. Achieving robust, scalable semantic integration of massive Next-Generation Sequencing (NGS) variant payloads within healthcare facilities presents a significant technical hurdle, as it requires balancing high-performance data management with the need to avoid exorbitant infrastructure costs and message bloat in operational Electronic Health Record (EHR) transaction systems. Current solutions have proven insufficient: standalone HL7 FHIR implementations suffer from extreme memory degradation when forced to handle bulk tabular variant lists,

while isolated Picture Archiving and Communication Systems (PACS) frameworks lack the necessary clinical semantic context. To resolve this, this study introduces a novel hybrid approach that shifts the paradigm by utilizing the pre-existing medical imaging PACS repository as a cost-effective, high-performance binary storage engine via custom registered IANA MIME-type encapsulation. By simultaneously deploying lightweight, targeted FHIR objects specifically for actionable clinical observations, this architecture establishes a clear separation of concerns between large-scale data archiving and structured diagnostic reporting, offering a dual capability that has been previously absent from unified systems.

Figure 1 maps out the high-level operational sequence. The workflow transitions sequentially from a state of slow, disjointed access toward automated ingestion, where data is centered within the institutional repository, processed with high-ratio compression algorithms for fast network retrieval, and organized via standardized DICOM metadata definitions to achieve immediate, unified clinical visualization.

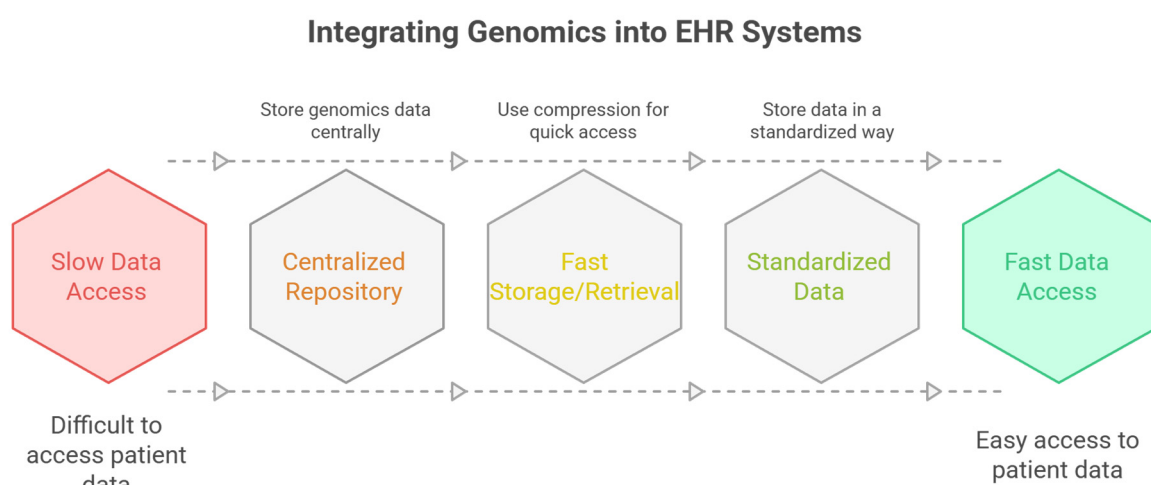


Figure 1. Integrating Genomics into EHR Systems.

2. DNA Data Compression

A common challenge in the field of DNA sequencing is the storage and management of its digital format. DNA sequencing generates massive amounts of data that need to be stored, managed, and analyzed efficiently and securely. One possible and efficient solution is to use the Picture Archiving and Communication System (PACS), which is widely used in medical imaging to store and retrieve modalities generated by imaging and signal data.

The most widely used format for representing the genomic variants called from a sequencing experiment is the Variant Call Format (VCF) [11], and it was developed for the 1000 Genomes Project [12], a large-scale project that aims to sequence and analyze the genomes of thousands of individuals from different populations. The VCF is a standard file format for storing and exchanging genomic variants, which are differences in the DNA sequence of an organism compared to a reference genome.

A VCF file is a text file that stores the variants found in a genome sequence, such as single nucleotide variants (SNVs), insertions, deletions, and structural variants.

In the context of hereditary heart diseases, these VCF files typically represent the output of either Whole Exome Sequencing (WES) or comprehensive cardiovascular gene panels (e.g., 100+ genes associated with cardiomyopathies and channelopathies). Because these diagnostic panels must be stored securely for long-term re-evaluation as pathogenic variant classifications evolve, efficient archiving is critical.

A VCF file consists of a meta-information block, a single header line and a data block. The meta-information lines, each prefixed by “##”, declare the format version, the reference genome, the filters applied and, through the ##INFO and ##FORMAT lines, the identifier, data type and cardinality of every

annotation field used in the file. The header line, prefixed by “#CHROM”, names the eight mandatory columns (CHROM, POS, ID, REF, ALT, QUAL, FILTER, INFO), followed by FORMAT and by one column per sample. The data block contains one tab-separated line per variant site. The standard organization of a VCF file, including its metadata, header, and data sections, is illustrated in Figure 2.

GENOMIC SEQUENCING: VCF FILE STRUCTURE (VCFv4.3)									
SECTION 1 - META-INFORMATION LINES									
<pre>##fileformat=VCFv4.3 ##fileDate=20230101 ##source=GenomicsPipeline ##reference=GRCh38 ##INFO=<ID=DP,Number=1,Type=Integer,Description='Total Depth'> ##INFO=<ID=AF,Number=A,Type=Float,Description='Allele Frequency'> ##INFO=<ID=MQ,Number=1,Type=Float,Description='RMS Mapping ##FILTER=<ID=LowQual,Description='Low quality'> ##FILTER=<ID=LowQual,Description='Low quality'> ##FORMAT=<ID=GT,Number=1,Type=String,Description='Genotype'> ##FORMAT=<ID=DP,Number=1,Type=Integer,Description='Read Depth'> ##FORMAT=<ID=GQ,Number=1,Type=Integer,Description='Genotype Quality'> ##contig=<ID=chr1,length=248956422></pre>									
SECTION 2 - HEADER LINE									
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT SAMPLE001 SAMPLE002									
SECTION 3 - DATA LINES									
chr1	10134	SNP123	A	T	99.0	PASS	GT:DP:GQ		
chr1	DP=45;AF=0.5;MQ=60.0						0/1:30:85	1/1:15:90	
chr1	10245	.	C	G,A	50.0	LowQual	GT:DP:GQ		
chr1	DP=20;AF=0.1,0.4;MQ=55.0						0/0:10:60	0/2:10:40	
chr1	10356	.	T		.	.	GT:DP:GQ		
chr1	DP=10;AF=0.2;MQ=40.0						0/0:5:40	0/1:5:35	
chr2	20567	SNP789	G	A	85.0	PASS	GT:DP:GQ		
chr2	DP=60;AF=0.95;MQ=60.0						1/1:40:95	1/1:20:99	

Figure 2. VCF structure. VCF is an open, community-maintained specification, currently developed under GA4GH. The figure is an original schematic prepared by the authors, illustrating the layout of a VCF file of the type produced by our sequencing workflow.

VCF files can be very large, especially when they include data from multiple individuals or populations, and take up a lot of disk space and computational resources. Therefore, compressing them can save disk space, reducing storage costs, bandwidth for transfers, and processing time. In addition to saving disk space and bandwidth, compressing VCF files can also improve the performance of downstream analyses. For example, compressed VCF files can be read more quickly than uncompressed files because they require less disk I/O.

A widely used datastore for large collections of variant calls is GenomicsDB [13], an open-source library, originally developed by Intel in collaboration with the Broad Institute, that stores variant data as sparse arrays. It is not a compressor in the sense used in this comparison: it is the storage layer invoked by GATK4 GenomicsDBImport to consolidate single-sample gVCFs prior to joint genotyping. It is therefore mentioned here for completeness and was not included in the benchmark of Table 1.

A free alternative to GenomicsDB is BCFtools [14], which is a set of command-line tools for manipulating VCF files. BCFtools can compress VCF files using the binary variant call format (BCF), which is a binary representation of the VCF format that uses zlib compression [15] and stores fields in a typed, variable-length binary encoding.

Table 1. Comparison of the compression of VCF files when compressed with each of the twelve compressors. It also shows the wall clock time average of each compression.

Compressor	Total Files	Avg Compression Time	Avg Compression Ratio
genozip	28	3.74	16.79
vcfshark	18	0.35	14.14
lzip	28	10.87	11.14
xz	28	13.35	11.10
bzip2	28	1.34	10.96
pbzip2	28	0.12	10.72
lzma	28	3.35	10.57
rar	28	0.30	9.18
zstd	28	0.21	8.93
gzip	28	0.35	8.59
zip	28	0.36	8.59
lz4	28	0.27	6.68

Another free solution for compressing VCF files is GenoType Compressor (GTC) [16], a tool designed for very large multi-sample genotype collections. GTC exploits the redundancy of the genotype matrix across samples by means of a Burrows–Wheeler-based permutation of the haplotypes followed by entropy coding, and can achieve higher compression ratios than BCFtools on such collections. Its advantage is, however, proportional to the number of samples jointly stored, and is therefore of limited relevance for the single-sample clinical VCFs considered in this work. A popular and highly efficient compressor for VCF files is VCFshark [17]. Its combination of high compression ratios, fast processing speeds, and low memory requirements makes it an excellent choice for managing large genomic datasets.

An open-source compressor like bzip2 [18] offers several advantages over other proprietary software, such as being free of charge for research or business use, security, and durability. Bzip2 is a compression algorithm that uses the Burrows–Wheeler method to transform the text into a sequence of repeated symbols, which can be encoded more efficiently.

Bzip2 is able to compress VCF files with an average ratio of 10:1 without loss of information. Moreover, being open-source software, bzip2 does not require the payment of maintenance and use licenses and guarantees the compatibility of the compressed files even after many years, unlike other software that could become obsolete or inaccessible. Bzip2 is therefore an ideal solution for compressing VCF files, saving space, time, and money. Genozip [19,20] is proprietary software for lossless compression of genomic files, including FASTQ, BAM, and VCF. Since it is designed specifically for genomic data, it is expected to perform better than general-purpose compressors. A parallel implementation of bzip2, pbzip2 [21], distributes the compression of independent blocks across multiple threads and was included in the comparison for this reason. The overall process of encapsulating VCF genomic data streams into standardized DICOM objects for PACS storage is depicted in Figure 3.

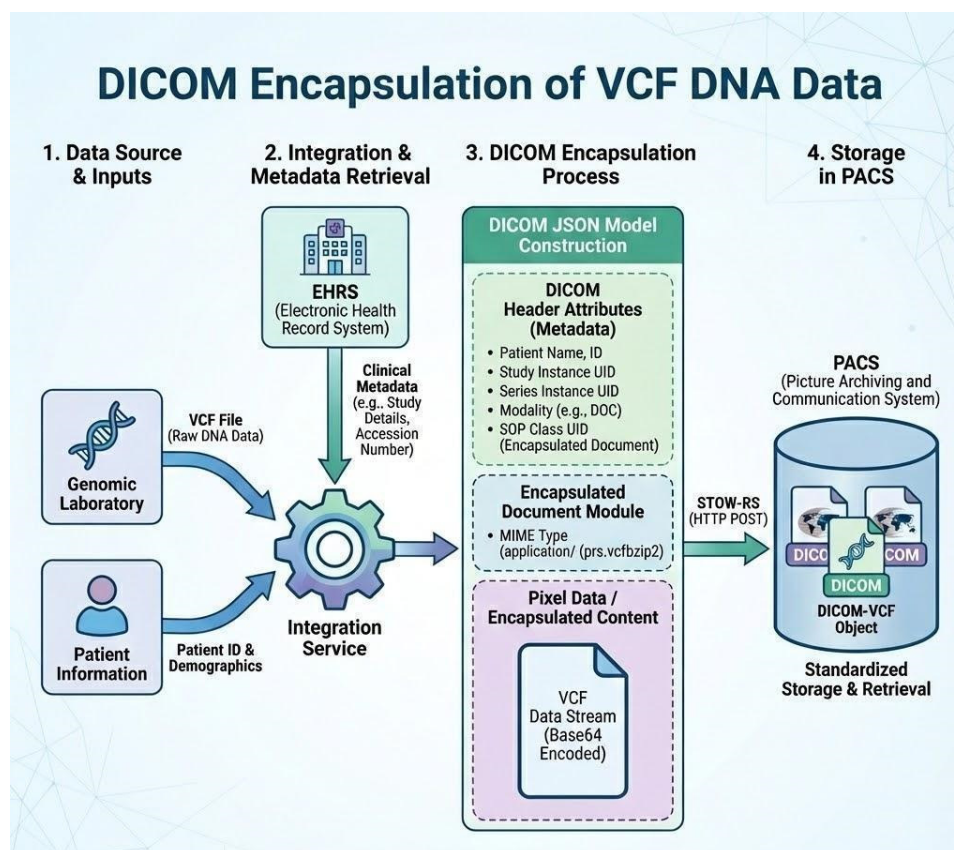


Figure 3. VCF DICOM Encapsulation.

3. Materials and Methods

The integration of genomic data from DNA sequencing into EHRs is a challenging task that requires novel solutions to be adopted on a large scale.

One of the challenges that hinders the access of DNA data from EHRs is the lack of complete interoperability with the sequencing equipment. The software that drives these analyzers is not designed to be compatible with EHRs, and therefore, it does not allow for a seamless transfer of information between the two systems. This creates difficulties in managing and interpreting the genomic data, as well as ensuring their quality and security. A possible solution would be to develop a standard interface that connects the sequencing equipment with the clinical records and that enables the exchange of DNA formats and demographic data in a consistent and reliable way.

A common example of an interface is the one defined by the DICOM standard, which enables the integration of diagnostic modalities, such as CT and MRI scans, with the EHRs. By associating the demographic data with the data produced by the imaging machine, the examination results and patient information can be easily retrieved while reducing human/typing errors and increasing safety. This is implemented using DICOM MWL (Modality Work List) and DICOM standard messaging. It would be possible to use DICOM standard messaging to retrieve the important information that characterizes the DNA data as belonging to the patient under study, but unfortunately, developing this type of standard messaging would necessitate access to the sequencer software source code or closer collaboration with the device manufacturer, currently not available for every producer, and hopefully achievable in the future.

Picture Archiving and Communication System (PACS) is used as the de facto standard for medical imaging and signal archiving. PACS are computer-based systems for storing, managing, and distributing medical images and data. They are used in hospitals, clinics, and other healthcare settings to increase the efficiency, accuracy, and accessibility of medical imaging.

The DICOM standard is the standard for medical imaging used in PACS. DICOM defines the format and structure of medical images and related data, and it also defines the protocols that are used to transmit medical images between different systems. Cardiology is a highly imaging- and signal-intensive specialty. Hospital PACS already natively store the patient's phenotypic manifestations (e.g., Echocardiograms, Cardiac MRIs, and ECGs). The core paradigm shift proposed in this methodology is utilizing the very same PACS infrastructure to archive the patient's genotype. This allows the cardiologist to retrieve and view the macroscopic structural/electrical phenotypes alongside the molecular genetic data (VCF/FHIR reports) within a single, unified clinical workspace.

One of the challenges of EHR integration is ensuring the security and privacy of sensitive patient information. A common approach is to use a secure application programming interface (API) to connect the EHRS with an external genomic database or laboratory.

However, this approach may not be feasible or efficient in some cases. In our scenario, the Illumina sequencer is connected exclusively to the hospital's internal network, and the data is stored locally, preventing any external integration. In such situations, a parallel approach may be more suitable. This approach involves using a web interface to the EHRS system to upload VCF files directly into the patient's record using their unique identifier (patient ID or other business identifier). This way, an internal or external laboratory can securely upload new genomic data to the EHRS without requiring an API or additional software.

The advantages of this parallel approach are its simplicity, velocity, and flexibility. It does not require any changes to the existing EHRS or to the genomic database.

The systemic barriers and technical challenges associated with standard EHR integration are visualized step-by-step in Figure 4. The diagram illustrates how core interoperability issues—stemming from a lack of native communication protocols between downstream sequencer equipment and hospital information layers—create significant data management difficulties. These obstacles require advanced mitigation strategies to resolve messaging format limitations and handle security and privacy concerns while maintaining direct network segregation.



Figure 4. Integrating Genomic Data into EHRs: Challenges and Solutions.

To overcome these structural limitations, a non-intrusive parallel integration approach was devised, as outlined in Figure 5. In this double-tier workflow, the standard clinical route for imaging and phenotypic clinical data entry runs in parallel with the genomic path. The sequencing laboratory uploads the compressed genomic binary payload directly to the secure internal integration service web interface while providing the required administrative patient identifiers. The integration node automatically creates a secure bridge, generating a standard Patient ID link that associates the un-truncated genomic file directly with the patient's master medical record. To eliminate human typing and transcription errors inherent to any parallel manual data entry, the integration service interface incorporates an automated validation layer. Upon entering the patient's business identifier, the service queries the hospital's central Master Patient Index (MPI). It retrieves and displays full demographic details (Name, Date of Birth, and Sex) for mandatory visual verification. Concurrently, a server script parses the uploaded VCF header to extract internal metadata identity tags, programmatically cross-checking them against the active EHR registry. If an identity conflict is detected, ingestion is immediately blocked to ensure clinical data integrity.

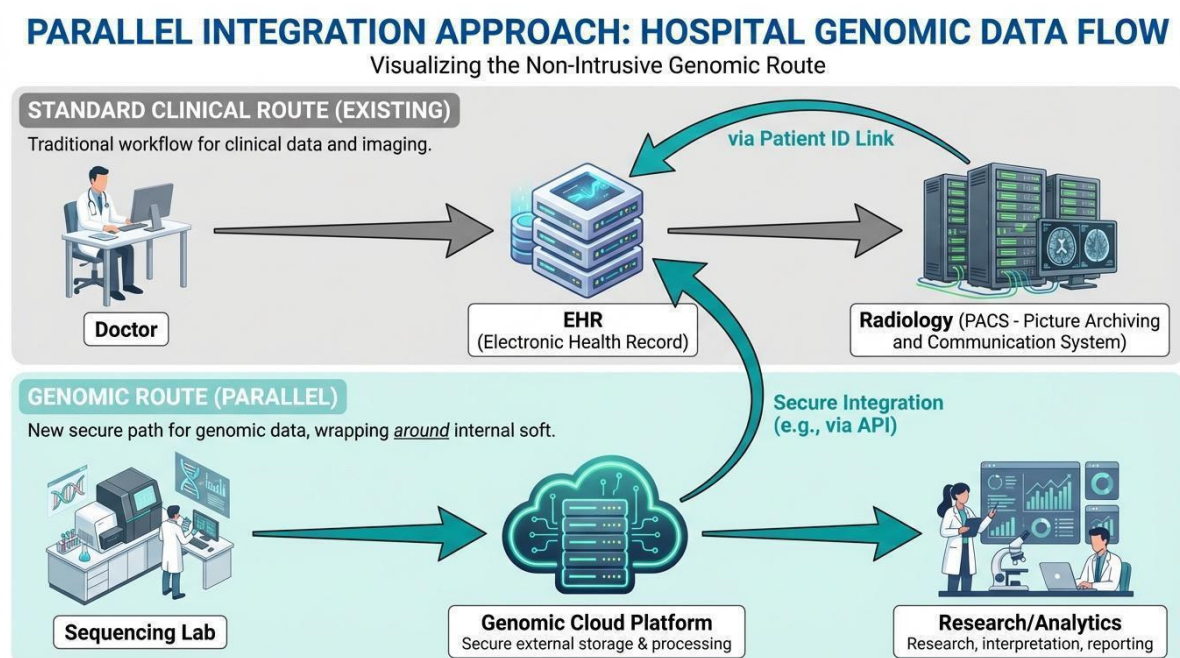


Figure 5. Parallel Integration Approach.

4. Results

In this study, we present the design and development of a framework and UX interface that streamlines the creation of DICOM files from VCF files. The system links genomic data with the patient's EHR and generates standardized DICOM objects for storage and access within a PACS environment.

The requirement for data compression was evaluated by comparing various algorithms applicable to Linux platforms. Our selection process prioritized identifying the most effective methods for DNA and VCF data, resulting in a comprehensive comparison of diverse toolsets and algorithmic performance.

The FTGM Molecular Cardiology Laboratory, specializing in hereditary heart disease—specifically cardiomyopathies and channelopathies—routinely conducts genetic analyses for approximately 600–700 patients annually. The VCF files generated by this high-volume clinical workflow served as the empirical dataset for evaluating long-term archiving and compression performance.

Our testing environment evaluated twelve compression tools against a dataset of VCF files produced by the Illumina Isaac Variant Caller (Starling2) v2.1.4.2, invoked by the DNA Enrichment Workflow module v3.0.0.23 of Illumina Local Run Manager on a NextSeq 550 platform (Illumina Inc., San Diego,

CA, USA) [22], using the UCSC hg19 assembly as reference. The matrix included widely available Linux command-line utilities and two domain-specific compressors: Genozip and VCFshark, both recognized for high-efficiency DNA data reduction.

The benchmarking results in Table 1 indicate that, while Genozip achieved the highest overall compression efficiency, general-purpose open-source tools like bzip2 and pbzip2 emerged as equally valid archival solutions. Although Genozip offers superior volumetric reduction, bzip2/pbzip2 provides a robust, license-free, and stable alternative for long-term archiving that avoids the risks of vendor lock-in. Pbzip2, which utilizes pthreads for parallel processing, demonstrated excellent speed, making it highly versatile in clinical settings. Conversely, VCFshark exhibited compatibility issues, failing to parse several clinical files, which underscores the risks of relying on highly specialized utilities that lack consistent maintenance. Ultimately, bzip2 and pbzip2 represent high-quality, dependable choices for institutions prioritizing open standards and multi-decade file accessibility.

To allow for technical replication, the experimental dataset utilized for architectural validation consisted of 28 distinct clinical VCF files (VCFv4.1, the version emitted by the Isaac Variant Caller) originating from targeted cardiovascular NGS sequencing runs at the FTGM Molecular Cardiology Laboratory. Individual uncompressed file payloads spanned a size distribution ranging from 45 MB up to 180 MB, containing germline variant profiles specific to cardiomyopathies and channelopathies panels. The computational architecture consisted of a localized Linux instance (CentOS Stream, 8-core CPU, 16GB RAM) hosting the integration service. The primary metrics tracked to evaluate the tool matrix included: (1) Compression Ratio, defined as the quotient of uncompressed bytes divided by compressed bytes; and (2) Wall Clock Execution Time, measuring the total elapsed time in seconds from execution command initiation to absolute file output closure on disk. System architectural protocols leveraged the DICOMwebTM standard framework, utilizing cURL-based multipart/related MIME streaming over HTTP POST to communicate directly with the open-source dcm4chee-arc-light core layer backend.

To ensure scientific objectivity and mitigate potential bias stemming from author affiliations with J4care GmbH and Genozip Limited, the benchmarking environment was strictly standardized using an automated Linux command-line script execution framework independent of proprietary tool interfaces. Tool selection criteria were strictly governed by practical clinical requirements: (1) long-term archival stability, ensuring format compliance over multi-decade clinical lifetimes; (2) decompression speed, which dictates clinical usability during urgent diagnostic sessions; (3) semantic interoperability within DICOM/HL7 healthcare networks; and (4) capital cost considerations.

While specialized genomic compressors like Genozip and VCFshark leverage domain-specific syntax structures to drastically reduce payload sizing, general-purpose open-source tools like bzip2/pbzip2 present a crucial risk-mitigation path. Open-source utilities offer foundational security guarantees through public code inspection and ensure file accessibility across arbitrary operating systems, without the vulnerabilities associated with commercial licensing. Conversely, proprietary software—despite offering superior volumetric reduction—carries vendor-lock risks, making a hybrid approach utilizing both tiers economically and operationally robust. Notably, Genozip offers free licenses for non-commercial academic research and open-source projects under specific conditions, which eases budget constraints in non-profit and research hospital environments.

The specific parsing failures encountered with vcfshark across 10 sample files were traced to strict syntax verification checks within its engine. The tool failed while parsing INFO and FORMAT fields that are specific to the Illumina Local Run Manager output and absent from the GATK-style VCFs against which the tool was developed, namely GQX, DPF and DPI in FORMAT, SNVSB, SNVHPOL, CIGAR, RU, REFREP and IDREP in INFO, and the TI, GI, EXON and FC annotation tags emitted by the MARS annotator, together with the non-standard meta-information lines ##annotator, ##content, ##SnpTheta, ##IndelTheta and ##VariantQualityScoringModel. Because vcfshark has not received recent updates to

accommodate these evolving diagnostic descriptor conventions, it encountered unrecoverable syntax rejections, emphasizing the clinical risks associated with deploying highly specialized utilities lacking ongoing maintenance support. The inclusion of the *VCFshark* utility in our benchmarking matrix was retained, as it is the sole domain-specific archiving alternative to the *Genozip* platform.

Operational validation of this framework was conducted within the live production ecosystem of the G. Monasterio Foundation (FTGM). Over the course of a multi-month clinical trial that mimicked continuous high-cadence sequencing services, key technical and usability performance metrics were analyzed. The average network time required for a clinical user to retrieve, download, and extract a complete VCF payload within the EHR interface via RESTful WADO-RS HTTP GET operations remained under 1.5 s for compressed bzip2 objects and under 2.2 s for Genozip objects, demonstrating excellent compliance with operational clinical timelines. From an administrative standpoint, the deployment of the automated gatekeeper validation interface effectively eliminated iterative human data transcription steps, resulting in an operational savings of approximately 15 min of administrative overhead per sequenced clinical batch. Longitudinal infrastructure monitoring confirmed that medical staff experienced zero slowdowns or latency spikes on traditional imaging PACS routines, as genomic objects are logically segregated under a unique “DNA” modality class registry, ensuring isolated database indexing.

Based on this balanced evaluation, both Genozip and bzip2/pbzip2 are presented as viable archival methodologies. The choice between them is driven by project-specific needs: institutions requiring maximum payload reduction may favor Genozip, while those prioritizing budget, long-term archival stability, and open standards will find bzip2/pbzip2 to be an optimal solution. Both paths ensure that genomic data is effectively managed within the clinical workflow. The sequential workflow for converting raw VCF files into DICOM-compliant objects is summarized in Figure 6.

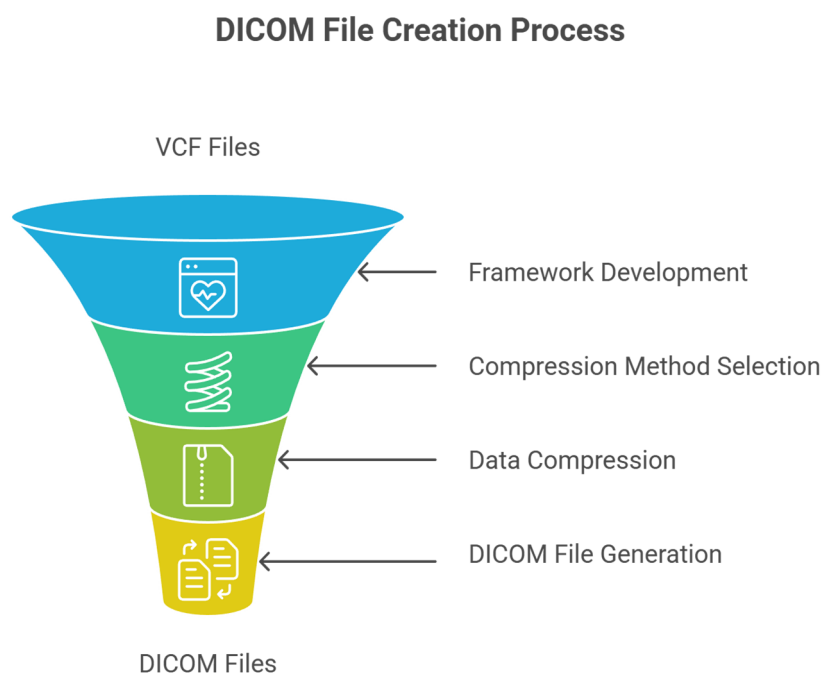


Figure 6. DICOM File Creation Process.

5. PACS Integration and EHR Workflow

A PACS allows the storage and management of DICOM-formatted objects. When it is necessary to store types of data that are not already defined in the DICOM standard, like VCF files and compressed VCF files, new capabilities need to be declared in order to support the extension, in an attribute called *dataType*, which identifies the format and content of the data.

The *data* type possible values are defined in the vocabulary Multipurpose Internet Mail Extensions (MIME) [23] by IANA (Internet Assigned Numbers Authority) [24], and in the version of the vocabulary, there were missing references for *genozip* and *bzip2*.

To comply with this constraint, we requested and registered two new entries in the IANA registry to define compressed VCF file *data* types:

- *prs.vcfbzip2* [25], *data* type value for *bzip2* compressed VCF files
- *Vnd.genozip* [26] *data* type value for *genozip* compressed VCF files

This registration enables official and standardized storage and retrieval of these files from a PACS without losing any information or DICOM standard compliance.

Within the DICOM standard architecture, Private SOP Classes are a fully compliant and standard extension mechanism intentionally engineered to allow healthcare institutions and equipment manufacturers to develop domain-specific functional extensions or integrate custom data objects before formal standardization committees define an official abstract syntax. According to standard DICOM conformance rules, when an external PACS receives an object containing a Private SOP Class that it does not natively implement, the receiving archive neither rejects the transaction nor generates structural syntax errors. Instead, the storage server accepts and archives the object as a generic encapsulated document, safely storing the underlying binary payload while preserving full database searchability via standard patient demographic metadata headers. Major imaging modality manufacturers (e.g., GE, Toshiba, Samsung) rely heavily on this extension protocol to maintain operational continuity across legacy infrastructures. Consequently, our implementation of custom UIDs ensures immediate backward compatibility and prevents file rejections across diverse multi-vendor hospital environments, meaning formal proposals to DICOM Working Group 30 are not operationally required at this phase.

The DICOM standard does not define a *SOP Class* (=Object Type) for the storage of VCF files, so two *private SOP Classes*, *private Dcm4che Encapsulated Genozip Storage* with *SOP Class UID* “1.2.40.0.13.1.5.1.4.1.1.104.1” and *Private Dcm4che Encapsulated Bzip2 VCF Storage* with *SOP Class UID* “1.2.40.0.13.1.5.1.4.1.1.104.2” were defined, following the *Information Object Definition* (IOD) for other *Encapsulated Documents* defined in DICOM PS 3.3 [27], with “*application/vnd.genozip*” and “*application/prs.vcfbzip2*” as values for DICOM attribute *MIME Type of Encapsulated Document* (0042,0012). Figure 7 shows an excerpt of the DICOM sequence dump displaying the custom MIME type tag set to *application/vnd.genozip*.

Number of Study Related Series	(0020,1206)	IS	1
Number of Study Related Instances	(0020,1208)	IS	1
Number of Series Related Instances	(0020,1209)	IS	1
MIME Type of Encapsulated Document	(0042,0012)	LO	application/vnd.genozip
Original Attributes Sequence	(0400,0561)	SQ	1 Item(s)
>	Item # 1		
> Modified Attributes Sequence	(0400,0550)	SQ	1 Item(s)

Figure 7. DICOM sequence dump excerpt, characterized by a specified MIME type *application/vnd.genozip*.

The DICOM attribute *Modality* (0008,0060) is set to “DNA” to facilitate searching the DICOM archive for studies that include such objects. An example of the DICOM header sequence dump featuring the custom ‘DNA’ Modality tag is presented in Figure 8.

Specific Character Set	(0008,0005)	CS	ISO_IR 100
Retrieve AE Title	(0008,0054)	AE	██████DNA
Instance Availability	(0008,0056)	CS	ONLINE
Modality	(0008,0060)	CS	DNA
Modalities in Study	(0008,0061)	CS	DNA
SOP Classes in Study	(0008,0062)	UI	1.2.40.0.13.1.5.1.4.1.1.104.1
Manufacturer	(0008,0070)	LO	ILLUMINA
Institution Name	(0008,0080)	LO	FTGM Massa
Manufacturer's Model Name	(0008,1090)	LO	ILLUMINA

Figure 8. DICOM sequence dump excerpt, characterized by “DNA” Modality, and Modalities in Study Tags.

To achieve the availability of DNA in EHRS, a web service interface was developed, capable of receiving from the genomic laboratory a VCF file and from the EHRS the patient’s identifier. To have a broader field of application for this integration service, the patient identifier was decided to be a unique business identifier (social insurance ID, fiscal identifier, *etc.*) assigned to every citizen and used to access various public services, including health care. Using the patient’s identifier, the integration service queries the EHRS. It retrieves the following information: Patient ID, issuer of Patient ID, Patient Name, study date, study time, patient birthdate, sex, and manufacturer, manufacturer model name, and modality were added in order to add Illumina DNA sequencer information. These details are essential to creating a proper DICOM file that conforms to the DICOM standard and contains the necessary metadata to identify the patient and the study.

In the EHRS, a specific action item related to a patient was developed that supports the creation and upload to the PACS of an encapsulated DICOM VCF, allowing the selection of a VCF file as a result of DNA sequencing for the selected patient.

The EHRS UX invokes the integration service with 2 parameters (Patient ID, compressed VCF file), and this creates a JSON file (Patient ID, Issuer of Patient ID, Patient Name, study date, study time, patient birthdate, sex, Manufacturer, Manufacturer model name, Modality) encoded according to the *DICOM JSON Model* specified in Annex F of DICOM PS 3.18 [28], and sends it as a multipart/related MIME message via an HTTP POST request according to the Store Transaction of the DICOMweb™ service (STOW-RS), also specified in DICOM PS 3.18 [29].

After positive acknowledgment, DNA retrieval is actionable for the clinical user via a button in the EHRS UX that launches the DNA retrieval and download service.

The EHRS can retrieve the DNA for the selected patient using the DICOM protocol: DICOM/DNA objects [30], which can be retrieved from the DICOM archive using DICOM Upper Layer Protocol Query/Retrieve Services specified in DICOM PS 3.4, Annex C [30], or by HTTP GET requests according to DICOMweb™ URI Services (WADO-URI) [30], or according to the Retrieve Transaction of the DICOMweb™ service (WADO-RS), also specified in DICOM PS 3.18 [31].

The two DICOMweb™ services provide options to extract the encapsulated compressed VCF file from the archived DICOM file by the server to return only the compressed VCF file, to be used by EHRs or other applications, without the capability to parse DICOM files.

As an example, we show in Figure 9 the Illumina BaseSpace Variant Interpreter, which is software that analyzes and interprets genetic variants derived from next-generation sequencing (NGS) data. It integrates bioinformatics tools and databases to provide annotation and filtering of variants, as well as evaluation of their potential clinical significance.

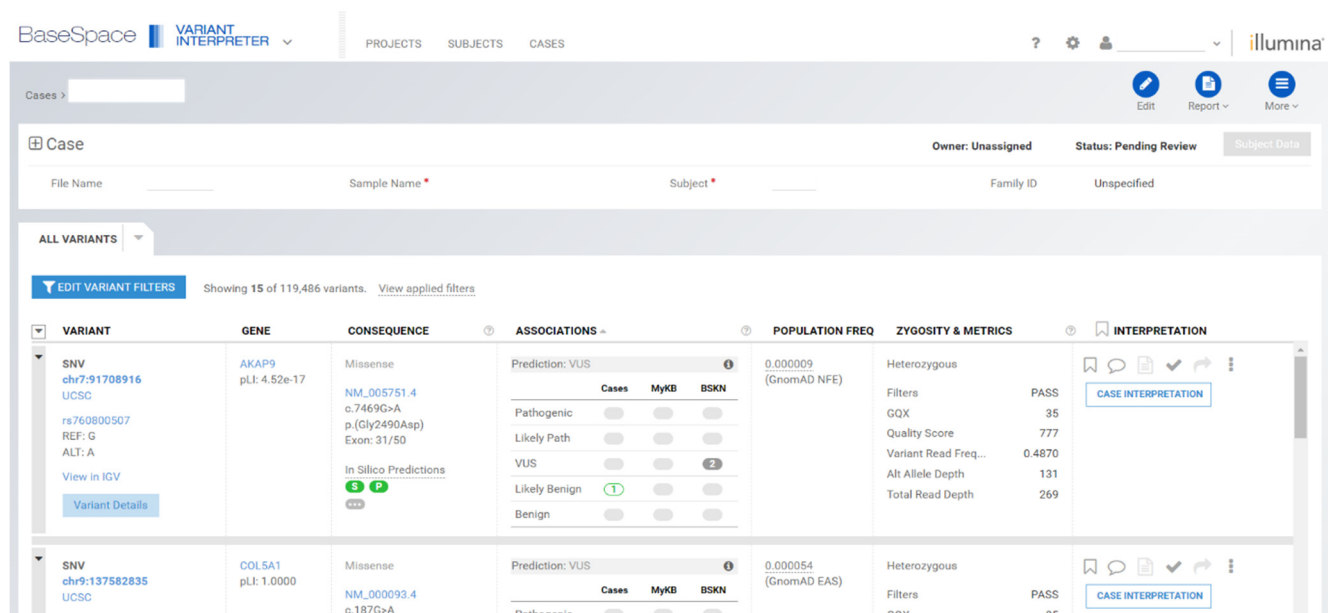


Figure 9. Illumina Dashboard.

The PACS used in the implementation was J4Care PACS [32] with a specialized software library called dcm4chee-arc-light [33], an open-source DICOM archive that can store and retrieve any kind of data as DICOM objects. The dcm4chee-arc-light was configured to accept the DNA VCF files as input and convert them into DICOM objects with a specific SOP class and transfer syntax.

Once the DNA VCF files were converted into DICOM objects, we could store them in the PACS J4care database and index them by their attributes. This allowed us to search for and retrieve them using the PACS J4care web interface.

By using this approach, we were able to integrate the DNA VCF files into the J4care PACS without modifying the core functionality of the application. We also preserved the original format and content of the DNA VCF files while making them accessible and interoperable with other DICOM systems.

On the EHR side, another possible solution explored was to use the HL7 FHIR standard to convert Variant Call Format (VCF) files into HL7 FHIR Genomics reports [34], which can then be exchanged and stored in EHRs. The utility *vcf2fhir* [35] implements this approach by creating FHIR resources for various genomic elements, such as variants, haplotypes, genotypes, annotations, and more. A major drawback of this method is the high memory consumption of these messages, which become unnecessarily long and bulky when stored in the EHR and require the EHR to handle efficiently and safely large amounts of data. A rigorous technical comparison reveals severe limitations in data preservation when relying solely on automated text translation tools like *vcf2fhir*. Due to the rigid structural discrepancies between tabular VCF schemas and hierarchical FHIR resource constraints, the *vcf2fhir* engine routinely strips away crucial primary data fields to prevent unmanageable text file expansion. Specifically, this converter systematically discards context-rich annotation metadata columns (e.g., VEP, SnpEff, dbSNP, ClinVar, and gnomAD annotations), complex structural or copy number variations (CNVs), multi-allelic site data, and

homozygous reference or no-call fields (./. or 0/0), preventing clinicians from verifying sequencing depth during subsequent diagnostic re-evaluations. Furthermore, the vcf2fhir engine enforces rigid parameters (such as the default ratio_ad_dp filter of 0.99 in vcf2fhir) that are highly restrictive, effectively filtering out legitimate heterozygous germline variants (which typically present an allele fraction of approximately 0.5) or low-frequency somatic variants, leading to silent, severe clinical data loss. Because precision cardiovascular medicine, and medicine in general, cannot tolerate data degradation—where rare germline variants are the predominant cause of hereditary channelopathies and copy number variants account for a clinically non-negligible minority of cases—the vcf2fhir converted file cannot be used as a patient’s valid genomic data.

Quantitatively, translating an entire clinical VCF file into individual FHIR resources generates immense message bloat, inflating simple JSON text strings to sizes between 50 MB and 150 MB, which degrades network broker performance. In contrast, our hybrid GACS model extracts and converts only actionable variants (typically fewer than ten clinically reportable variants per patient) into structured FHIR Observation resources, dropping the transactional EHR payload to less than 15 KB. The complete, un-truncated genomic dataset is safely archived as a compressed binary file within the PACS layer and accessed via a lazy-loading protocol using a secure WADO-RS URL embedded in the FHIR DocumentReference resource. This hybrid model achieves an immediate data reduction of approximately 99.9% on active hospital network transaction loops while guaranteeing that the complete variant call set is archived without truncation.

A second common obstacle is related to the IT implementation: most hospitals currently do not adopt the HL7 FHIR architecture for their internal systems, but only for external communication and integration purposes. Therefore, this approach is currently not compatible with the existing infrastructure and standards of many health care providers, but, given the evolution of communication standards toward HL7 FHIR, it could represent a future domain for analysis and experimental use. The end-to-end operational process flow between the Illumina sequencer, integration service, PACS archive, and the clinical workstation is illustrated in Figure 10.

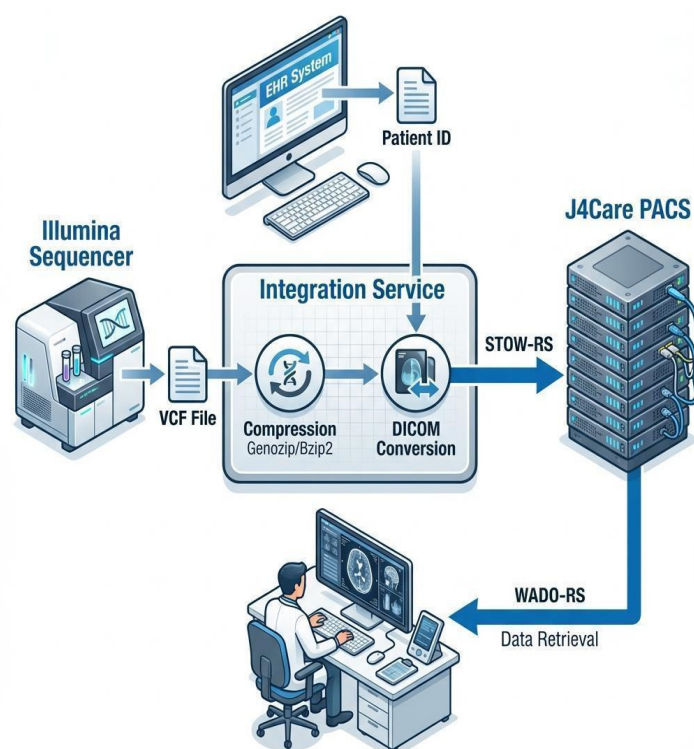


Figure 10. DNA and PACS Process Flowchart.

6. Future Directions

PACS as a Genomic Archiving and Communication System (GACS) with HL7 FHIR

While the presented DICOM encapsulation method successfully addresses the storage and standardized retrieval of VCF files, a key limitation remains in semantic interoperability with modern Electronic Health Record (EHR) systems that increasingly adopt the HL7 FHIR standard. The previously discussed drawback of native FHIR implementation was the high memory consumption of bulk genomic reports.

To bridge this gap, the PACS can be conceptually extended to function as a Genomic Archiving and Communication System (GACS), supporting a hybrid integration architecture. This model treats the PACS not just as a bulk data archive, but as a universal repository where raw VCF data (stored via DICOM Encapsulation) is linked to structured clinical reports represented by FHIR resources.

The core of this architecture relies on two complementary systems:

- **PACS Data Layer (DICOM/GACS):** The existing PACS archive continues to store the highly compressed VCF files (using methods like Genozip or bzip2, but not limited to) as immutable DICOM Encapsulated Documents. It leverages its existing indexing capabilities (Modality = “DNA”) and the WADO-RS (Web Access to DICOM Objects—RESTful Services) interface to expose this data. WADO-RS acts as a secure, firewall-friendly delivery layer for raw/compressed genomic files.
- **EHR/Clinical Layer (FHIR):** A FHIR server within the EHR environment stores structured clinical information. Key FHIR resources are used for clinical decision support.
- **DiagnosticReport:** Synthesizes the clinical findings derived from the genomic data.
- **Observation (Variant Details/Implications):** Captures individual clinically significant variants and allelic depths to enable clinical decision support, particularly for resulting cardiac diagnostics (e.g., hypertrophic cardiomyopathy) and related therapeutic implications. Most notably, a mature HL7 FHIR integration with Clinical Decision Support (CDS) systems enables automated genotype-informed prescribing alerts. For example, the EHR could raise an alert at the point of prescription of QT-prolonging medications in patients carrying a pathogenic or likely pathogenic variant in a Long QT Syndrome gene, cross-referenced against a curated database of drugs with QT liability. Such an alert should remain non-interruptive and overridable: a negative genetic test does not exclude clinically diagnosed Long QT Syndrome, and QT-prolonging drugs may still be indicated after individual risk assessment.
- **DocumentReference:** This crucial resource links the FHIR objects to the raw VCF file in the PACS/GACS. It contains a WADO-RS URL, allowing a clinician or an automated analysis tool to retrieve the compressed VCF data on demand (lazy loading). The proposed model for linking FHIR DiagnosticReport and Observation resources with PACS data via WADO-RS links is shown in Figure 11.

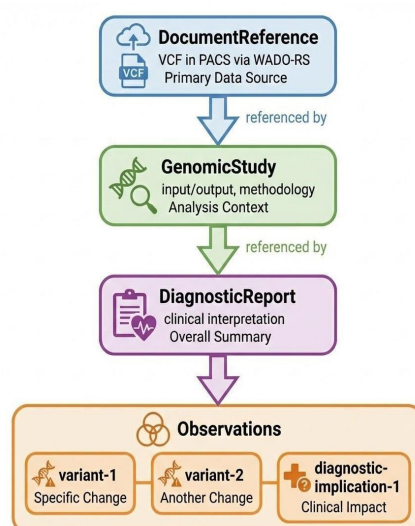


Figure 11. Proposed Model for Diagnostic Report using wado.

To manage the challenge of large file sizes, this hybrid strategy converts only the clinically significant genetic variants into FHIR observations. Simultaneously, it maintains a secure, standards-compliant link to the complete, compressed VCF file. This full file is stored in a high-performance PACS/GACS archive, making it readily available for auditing or comprehensive bioinformatics analysis.

7. Conclusions

The integration of DNA sequencing data into existing healthcare systems represents a significant advancement for precision medicine. By linking genetic data with cardiovascular imaging and ECG signals, clinicians can facilitate seamless data exchange across clinical networks. This study demonstrates that both proprietary and open-source compression methods are capable of supporting this evolution. While this integration is transformative, it necessitates a choice of archival strategy driven by specific institutional requirements—whether prioritizing the high-ratio efficiency of specialized tools or the dependable, license-free stability of open standards. Future research should continue to explore standardized protocols, such as FHIR Genomics, in conjunction with these PACS/GACS models to further enhance clinical decision support.

Regarding the performance impact on clinical workflows and network infrastructure, a comparative analysis indicates that the integration of genomic data does not compromise system efficiency. For context, an average cardiology PACS CT study in our environment consists of more than 1–20 thousand images, amounting to a data volume of 1–10 GB, and can take up to 5 min to complete downloading over the network on busy days. In comparison, a compressed 180 MB VCF file is transferred very rapidly. Within our operational setting, the payload of a VCF file is highly comparable to a standard ultrasound (US) study, which typically ranges between 20 MB and 200 MB. Consequently, VCF files represent an average-sized study profile that is well tolerated and seamlessly accommodated within existing PACS environments without causing degradation in standard imaging performance.

7.1. Technical and Operational Solutions

To overcome practical challenges, particularly the native non-compliance of the VCF format with the DICOM standard, using DICOM encapsulated document objects or private tags is recommended. The management of large genomic files is mitigated by leveraging WADO-RS with multipart support and compression (e.g., gzip/bzip2). To enable genomic search and analysis in a clinical context, implementing FHIR Genomics Operations is crucial for on-demand reading of VCF data from the PACS. A hybrid architecture, where PACS serves as the *storage* for raw data and a FHIR server *caches* the results of VCF to FHIR conversions, is suggested for optimal performance. Security and privacy must be maintained through a combination of DICOM encryption, FHIR authorization mechanisms, and a dual *audit logging* system. The feasibility of archiving non-DICOM formats like VCF is demonstrated by the capability of modern PACS to support *custom content-types* in STOW-RS.

7.2. Economic Advantages

This integrated PACS-FHIR approach offers significant economic benefits compared to implementing a separate Genomics Archiving and Communication System (GACS). The technical implementation details and economic advantages of the integrated PACS-FHIR model compared to a standalone GACS infrastructure are outlined in Table 2.

The path forward requires a continuous focus on standardization (especially the FHIR Genomics Reporting Standard), ensuring interoperability, and adopting robust protocols that balance data accessibility with security and scalability needs, charting the course towards a truly personalized and data-driven healthcare system.

Table 2. Advantages of the integrated PACS-FHIR approach.

Advantage	Technical Implementation	Benefit
Cost Avoidance	No separate GACS purchase/license required.	Saves an estimated €100K–€500K in GACS infrastructure costs.
Storage Savings	Utilizing compression (e.g., Genozip) with existing PACS storage.	Achieves immense data volume reduction, potentially avoiding petabytes of storage.
Reduced Licensing	Leveraging open-source tools (e.g., HAPI FHIR, dcm4chee).	Minimizes proprietary software licensing costs.
Operational Efficiency	Single IT team manages the unified PACS-FHIR system.	Reduces overhead in training, support, and maintenance compared to managing separate systems.
Scalability	PACS storage scales linearly; FHIR queries are cached.	Decreases the marginal cost per additional patient over time.

While specialized genomic compressors like Genozip leverage domain-specific structures for superior volumetric reduction, our selection was primarily driven by the need for institutional reliability. Genozip is an enterprise-grade product widely deployed in petabyte-scale clinical repositories, ensuring professional maintenance and active development. In contrast, VCFshark functions as a very good proof of concept; its lack of formal customer support and regular maintenance releases limits its suitability for production-grade clinical environments. Consequently, we pair Genozip with bzip2/pbzip2 not only as a cost-based compromise but also as a strategic risk-mitigation layer, ensuring long-term data accessibility through open-source standards independent of future proprietary support cycles.

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

Throughout the development of this manuscript, the authors leveraged generative artificial intelligence and advanced AI-assisted large language model frameworks (Gemini 3.0–3.5) exclusively to optimize linguistic articulation, strengthen structural coherence, and refine formal academic prose, as well as to generate high-resolution illustrative assets for various diagrams and graphical figures. These tools operated under strict human oversight and were not employed to synthesize original empirical data, execute statistical benchmarking, or formulate core scientific and methodological conclusions. The authors remain fully responsible and accountable for the integrity, veracity, and accuracy of all scientific statements and technical contents presented in this study.

Acknowledgments

The authors would like to express their sincere gratitude to Alessio Ciregia (FTGM) for his valuable technical and administrative support. Special thanks are also extended to Petr Kalina, Gottfried Jäger, Josef Friedrichkeit, and Michael Knapp (J4care) for their technical expertise and assistance throughout the project. Additionally, the authors acknowledge Illumina for providing technical support regarding the Variant Call Format (VCF) files and Illumina device.

Author Contributions

Conceptualization, P.M., S.D. and M.S.; Methodology, P.M., S.D., G.Z. and D.M.L.; Software, P.M., G.Z., V.N. and D.M.L.; Validation, P.M., S.D., M.P., M.M., S.V. and N.B.; Formal Analysis, P.M., M.P., M.M. and D.M.L.; Investigation, P.M., S.D., M.P., M.M., S.V. and N.B.; Resources, S.D., N.B. and M.S.; Data Curation, P.M., M.P., M.M.; Writing—Original Draft Preparation, P.M., M.M. and M.P.; Writing—Review & Editing, P.M., S.D., M.P., M.M., G.Z., V.N., D.M.L., S.V., N.B. and M.S.; Visualization, P.M.

and M.M.; Supervision, S.D. and M.S.; Project Administration, P.M. and S.D.; Funding Acquisition, S.D. and M.S.

Ethics Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the G. Monasterio Foundation (FTGM) under authorization Deliberation n. 499 of 28 September 2026. The need for individual ethical approval was waived for this technical study because it involved the retrospective analysis of existing clinical data, which were managed and processed in full compliance with the privacy policies and the General Data Protection Regulation (GDPR) of the institution. No experimental interventions on patients were performed.

Informed Consent Statement

Patient consent for this specific technical study was waived by the Institutional Review Board, given that the research involved the retrospective utilization of anonymized/pseudonymized clinical data generated during standard diagnostic procedures. All patient data were processed in accordance with the G. Monasterio Foundation's internal protocols for secondary use of health data for research purposes.

Data Availability Statement

The data presented in this study are available on request from the corresponding author, subject to institutional approval and privacy constraints. The raw genomic data are protected clinical records and cannot be made publicly available in accordance with GDPR and institutional privacy policies.

Funding

This research was supported by institutional research funds from the Fondazione G. Monasterio (FTGM) and J4care GmbH, under the scientific leadership and supervision of Stefano Dalmiani (FTGM) and Marcel Swennenhuis (J4care GmbH).

Declaration of Competing Interest

The authors disclose the following potential competing interests: Gunter Zeilinger, Vrinda Nayak, and Marcel Swennenhuis are affiliated with J4care GmbH, which is the vendor for the J4CARE VNA PACS archive used and discussed in the implementation (based on the open-source DCM4CHEE Archive 5.x). Divon Mordechai Lan is affiliated with Genozip Limited, a company that provides the proprietary Genozip compression software, which was selected as the first-choice compressor in this study due to its superior compression ratio.

References

1. Kullo IJ, Jarvik GP, Manolio TA, Williams MS, Roden DM. Leveraging the electronic health record to implement genomic medicine. *Genet. Med.* **2013**, *15*, 270–271. DOI:10.1038/gim.2012.131
2. Nishimura AA, Tarczy-Hornoch P, Shirts BH. Pragmatic and Ethical Challenges of Incorporating the Genome into the Electronic Health Record. *Curr. Genet. Med. Rep.* **2014**, *2*, 201–211. DOI:10.1007/s40142-014-0051-9
3. Schmidt S, Sax U. Integration of Genomic Data in Electronic Health Records. *Methods Inf. Med.* **2005**, *44*, 546–550. DOI:10.1055/s-0038-1634006
4. Shoenbill K, Fost N, Tachinardi U, Mendonca EA. Genetic data and electronic health records: A discussion of ethical, logistical and technological considerations. *J. Am. Med. Inform. Assoc.* **2014**, *21*, 171–180. DOI:10.1136/amiajnl-2013-001694
5. Goodwin S, McPherson JD, McCombie WR. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. DOI:10.1038/nrg.2016.49

6. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, et al. Fluorescence detection in automated DNA sequence analysis. *Nature* **1986**, 321, 674–679. DOI:10.1038/321674a0
7. Satam H, Joshi K, Mangrolia U, Waghoo S, Zaidi G, Rawool S, et al. Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology* **2023**, 12, 997. DOI:10.3390/biology12070997
8. Carter AB, Abruzzo LV, Hirschhorn JW, Jones D, Jordan DC, Nassiri M, et al. Electronic Health Records and Genomics: Perspectives from the Association for Molecular Pathology Electronic Health Record (EHR) Interoperability for Clinical Genomics Data Working Group. *J. Mol. Diagn.* **2022**, 24, 1–17. DOI:10.1016/j.jmoldx.2021.09.009
9. Grebe TA, Khushf G, Chen M, Bailey D, Brenman LM, Williams MS, et al. The interface of genomic information with the electronic health record: A points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* **2020**, 22, 1431–1436. DOI:10.1038/s41436-020-0841-2
10. Son JH, Xie G, Yuan C, Ena L, Li Z, Goldstein A, et al. Deep Phenotyping on Electronic Health Records Facilitates Genetic Diagnosis by Clinical Exomes. *Am. J. Hum. Genet.* **2018**, 103, 58–73. DOI:10.1016/j.ajhg.2018.05.010
11. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* **2011**, 27, 2156–2158. DOI:10.1093/bioinformatics/btr330
12. 1000 Genomes | A Deep Catalog of Human Genetic Variation. Available online: <https://www.internationalgenome.org/> (accessed on 9 August 2023).
13. Datta K, Gururaj K, Naik M, Narvaez P, Rutar M. GenomicsDB: Storing Genome Data as Sparse Columnar Arrays. Available online: <https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/genomics-storing-genome-data-paper.pdf> (accessed on 12 August 2026).
14. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, 10, giab008. DOI:10.1093/gigascience/giab008
15. zlib Home Site. Available online: <https://zlib.net/> (accessed on 13 June 2024).
16. Danek A, Deorowicz S. GTC: How to maintain huge genotype collections in a compressed form. *Bioinformatics* **2018**, 34, 1834–1840. DOI:10.1093/bioinformatics/bty023
17. Deorowicz S, Danek A, Kokot M. VCFShark: How to squeeze a VCF file. *Bioinformatics* **2021**, 37, 3358–3360. DOI:10.1093/bioinformatics/btab211
18. McCool M, Robison AD, Reinders J. *Bzip2 Data Compression*; Elsevier: Amsterdam, The Netherlands, 2012; pp. 291–297. DOI:10.1016/B978-0-12-415993-8.00012-8
19. Lan D, Tobler R, Souilmi Y, Llamas B. genozip: A fast and efficient compression tool for VCF files. *Bioinformatics* **2020**, 36, 4091–4092. DOI:10.1093/bioinformatics/btaa290
20. Lan D, Tobler R, Souilmi Y, Llamas B. Genozip: A universal extensible genomic data compressor. *Bioinformatics* **2021**, 37, 2225–2230. DOI:10.1093/bioinformatics/btab102
21. Parallel BZIP2 (PBZIP2). Available online: <http://compression.great-site.net/pbzip2/> (accessed on 14 June 2024).
22. Illumina | Sequencing and Array-Based Solutions for Genetic Research. Available online: <https://www.illumina.com/> (accessed on 9 August 2023).
23. Freed N, Borenstein N. *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*; Request for Comments RFC 2045; Internet Engineering Task Force: Wilmington, DE, USA, 1996. DOI:10.17487/rfc2045
24. Internet Assigned Numbers Authority. Available online: <https://www.iana.org/> (accessed on 14 June 2024).
25. VCFBZIP2 Iana Assignment. Available online: <https://www.iana.org/assignments/media-types/application/prs.vcfbzip2> (accessed on 6 September 2023).
26. Genozip IANA Assignment. Available online: <https://www.iana.org/assignments/media-types/application/vnd.genozip> (accessed on 6 September 2023).
27. A.45 Encapsulated Document IOD. Available online: https://dicom.nema.org/medical/dicom/current/output/cthtml/part03/sect_A.45.html (accessed on 12 June 2024).
28. PS3.18. Available online: https://dicom.nema.org/medical/dicom/current/output/html/part18.html#chapter_F (accessed on 12 June 2024).
29. 10.5 Store Transaction. Available online: https://dicom.nema.org/medical/dicom/current/output/cthtml/part18/sect_10.5.html (accessed on 12 June 2024).
30. PS3.4. Available online: https://dicom.nema.org/medical/dicom/current/output/html/part04.html#chapter_C (accessed on 12 June 2024).
31. Overview-FHIR v5.0.0. Available online: <https://www.hl7.org/fhir/overview.html> (accessed on 12 June 2024).
32. J4Care-Home of DCM4CHEE, J4Care-Home of DCM4CHEE. Available online: <https://www.j4care.com/> (accessed on 14 June 2024).

33. Zeilinger G. dcm4che/dcm4chee-arc-light: DICOM Archive J2EE application. Available online: <https://github.com/dcm4che/dcm4chee-arc-light> (accessed on 16 May 2024).
34. HL7.FHIR.UV.GENOMICS-REPORTING\Home Page-FHIR v4.0.1. Available online: <https://www.hl7.org/fhir/uv/genomics-reporting/index.html> (accessed on 14 June 2024).
35. Dolin RH, Gothi SR, Boxwala A, Heale BSE, Husami A, Jones J, et al. vcf2fhir: A utility to convert VCF files into HL7 FHIR format for genomics-EHR integration. *BMC Bioinform.* **2021**, *22*, 104. DOI:10.1186/s12859-021-04039-1