

Cloud computing-based collaborations on health research: A systematic review of applications, barriers to adoption, and migration strategies

Robert Mugonza^{1,*}, Richard Ssembatya¹, Annabella Basaza-Ejiri²

ABSTRACT

Background: Modern biomedical research is undergoing a fundamental transformation driven by an exponential increase in data complexity and the integration of artificial intelligence and cloud computing into biomedical and public health infrastructure. While cloud computing offers a pivotal solution to the limitations of on-premise infrastructure, the transition is hindered by a complex set of domain-specific challenges. This systematic review maps the current landscape of cloud adoption, identifies systemic barriers to wider usage, and outlines practical requirements for successful cloud migration.

Methods: We conducted a comprehensive systematic search of articles published between 2010 and early 2026 across PubMed and Google Scholar. To capture both clinical and technical literature, the search was expanded to include major computer science and engineering databases, namely IEEE Xplore and Web of Science. We identified 68 primary studies evaluating 50 unique applications and subjected these to a formal assessment using a 5-point informatics quality appraisal (IQA) scale to determine the maturity of evidence.

Results: Our analysis revealed a highly mature ecosystem, with 96% of identified applications reaching production-ready or validated implementation stages (IQA scores = 3-5). These applications were categorized into three functional types: task-specific tools, integrated platforms, and large-scale archives. Despite a projected market growth exceeding 17%, however, a significant readiness gap persists. Innovation remains concentrated in North America and is characterized by high data-egress costs, regulatory misalignments, and friction associated with technical interoperability. In contrast, initiatives such as the District Health Information System-2 (DHIS-2) demonstrate the potential for cloud applications to drive global health equity in resource-limited settings.

Conclusion: The cloud has successfully moved from being a conceptual framework to serving as the backbone of modern health research. Nevertheless, a considerable trust deficit coupled with persistent structural barriers related to costs, data sovereignty, and regulatory complexity continue to impede global collaboration. To fully realize the promise of democratizing research, future development must transition from performance-centric models to federated, context-aware architectures that prioritize accessibility, inclusivity and secure data sharing. From a policy and practice perspective, funding agencies, governments, and international research consortia should support the adoption of such architectures while promoting harmonized cross-border data governance frameworks and measures to reduce prohibitive cloud data-egress costs. These efforts are essential for enabling equitable participation in global health research ecosystems, particularly among institutions in low- and middle-income countries.

Key words: cloud computing, cloud migration, data security, health research, systematic review

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

¹Department of Computer Science, Faculty of Computing and Informatics, Mbarara University of Science and Technology, P.O. Box 1410, Mbarara, Uganda, ²Department of Information Technology, College of Computing and Information Sciences, Makerere University, P.O. Box 7062, Kampala, Uganda.

***Address for correspondence:** Department of Computer Science, Faculty of Computing and Informatics, Mbarara University of Science and Technology, P.O. Box 1410, Mbarara, Uganda. Email: robertmugonza@must.ac.ug; <https://orcid.org/0000-0001-8121-629X>

Copyright © 2026 The Authors. Published by Wolters Kluwer Health, Inc. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0, where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially.

Digit Med 2026;12:e26-00006

Received: 4 February 2026; **Accepted:** 9 May 2026

<https://doi.org/10.1097/DM-2026-00006>

INTRODUCTION

Modern biomedical and public health research is experiencing profound change prompted by an exponential increase in data volume and complexity as well as the recent integration of liberating technologies, including cloud computing and artificial intelligence (AI), into scholarly endeavors.^[1,2] In this field, effectively leveraging diverse datasets, from those encompassing high-throughput genomics to extensive electronic health records and real-time public health surveillance, demands large-scale interdisciplinary collaborations.^[3,4] The problem is that traditional on-premise computing infrastructures often fall short, constrained by high acquisition costs, limited scalability, and complex data exchange processes. These issues, in turn, hinder the agility required for data-intensive research.

Yet the abovementioned challenges can be resolved by the pivotal technology that is cloud computing, whose influence has tremendously benefited the establishment of databases and archives for diseases of interest, effective data processing,^[5] and the development of applications hosted on advanced computing infrastructures, such as biomedical^[6] and e-health^[7] clouds. Cloud platforms provide on-demand access to scalable computational resources, flexible storage, and advanced analytical tools, uniquely positioning them to support the rigorous demands of contemporary health research. The inherent affordances of cloud services, such as ubiquitous access, resource pooling, multitenancy, and metered consumption,^[8] facilitate robust resource utilization and seamless data sharing,^[9] thereby enabling the dissemination of application services (data and computation) and infrastructure by organizations, companies, and individuals in a robust and more seamless manner.^[8,10] This paradigm shift has also enabled researchers to transcend local infrastructural shortcomings, dynamically scale resources, and dedicate more focus to scientific inquiry and algorithmic advancement.

The cloud's capabilities are extended by most of the infrastructure required for biomedical and health research today, such as Basic Local Alignment Search Tool (BLAST).^[11] This advantage has, in turn, significantly enhanced data access and management, providing new avenues in which to derive insights into disease understanding and patient care.^[5,12-14] These remarkable benefits and increasing adoption are evidenced by major government and institutional initiatives, such as AWS GovCloud, the NIH's STRIDES, and the European Open Science Cloud.^[15] Nevertheless, the transition is hindered by a complex set of domain-specific challenges. For example, while cloud integration is promising, it introduces formidable obstacles regarding security,^[16,17] privacy,^[5,18,19] interoperability,^[20,21] and regulatory compliance.^[19] Additionally, although some of these challenges are presented in related works, no comprehensive and systematic analysis has been devoted to the specific impediments to the effective use of cloud platforms for collaborative biomedical and public health research.

Correspondingly, this review addressed the critical gaps described above by surveying the application of cloud computing in collaborative biomedical and health research, with specific focus directed toward the data sharing approaches employed and the obstacles encountered in this discipline. The last two decades have witnessed the rapid evolution of the cloud landscape, prompting us to explore the current state of adoption in the domain of interest, identify exemplary cloud computing applications, document the adoption of such technologies, and ascertain the practical requirements for migrating research collaborations and their data onto cloud platforms. Our review was guided by the following research questions (RQs). RQ1: What is the current state of cloud computing adoption, and what are the exemplary cloud applications used in biomedical and public health research? RQ2: What are the specific opportunities and major challenges presented by these applications? RQ3: On the basis of the current landscape and the identified challenges, what practical requirements and strategies are necessary for successfully migrating biomedical and public health research collaborations onto cloud platforms?

By addressing these questions through a comprehensive systematic review, this study establishes a necessary benchmark for cloud-integrated health informatics. Moving beyond a

descriptive catalog of technologies, this work translates complex technical and regulatory barriers into structured requirements for cloud migration. Ultimately, this framework aims to provide researchers, cloud architects, and international funding bodies with the strategic clarity needed to deploy equitable, secure, and cross-border collaborative ecosystems).

The remainder of this article is organized as follows: Section 2 describes the methodology adopted in the review, while Section 3 presents the data collected to address RQ1 and the findings related to RQ2. The presentation focuses on the application types observed and the adoption levels reported in the examined studies. Section 4 discusses the results, details the major barriers to adoption, and clarifies the constraints on wider usage. Section 5 describes the strategic guidance that we formulated, addressing RQ3, and Section 6 concludes the paper with the limitations of the study and implications for research, practice, and policymaking.

METHODS

We conducted a systematic literature review and synthesized qualitative evidence, with the procedures and results reported in accordance with the 2020 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework.^[22] While the PRISMA 2020 guidelines encompass meta-analyses, this review focused on a qualitative synthesis to map the current landscape of cloud applications, identify key opportunities, and discern recurring challenges. The PRISMA methodology was employed to ensure transparency, completeness, and reproducibility. Because our research objective was to categorize infrastructure, tools, and systemic barriers rather than pool statistical effect sizes, a narrative and thematic synthesis was deemed the most appropriate method for data integration. The approaches used are detailed in the succeeding subsections.

Literature search and strategy

We conducted a systematic search across both biomedical and technical research databases, namely, PubMed, Google Scholar, IEEE Xplore, and Web of Science, targeting studies on the application of cloud computing in biomedical and health research collaborations published from January 2010 to early 2026. This period captures the technological transition from the initial adoption of cloud technology to the current phase characterized by its use as support for AI-driven health discovery. The initial survey (concluded in October 2025) focused on PubMed and Google Scholar, but the search was expanded in February 2026 to include IEEE Xplore and Web of Science. This update ensured the inclusion of the technical architectures and computer science proceedings essential to the cloud computing domain and the coverage of the most recent peer-reviewed advancements. The survey encompassed English-language articles from journals and conference proceedings, and search terms were combined using Boolean operators (AND, OR) to refine the strategy. The specific search syntaxes are listed below.

(1) Google Scholar - ("cloud computing" OR "cloud-based") AND ("health research" OR "biomedical research") AND ("adoption" OR "barriers" OR "migration" OR "strategy") AND ("collaboration" OR "consortia")

(2) PubMed - ("cloud computing" [tiab] OR "cloud based"

[tiab]) AND ("health research" [tiab] OR "biomedical research" [tiab] OR "healthcare" [tiab]) AND ("adoption" [tiab] OR "barriers" [tiab] OR "migration" [tiab] OR "strategy" [tiab] OR "implementation" [tiab]) AND ("collaboration" [tiab] OR "consortia" [tiab])

(3) Web of Science: TS = (("cloud computing" OR "cloud-based") AND ("biomedical research" OR "health research") AND ("collaboration" OR "consortia") AND ("adoption" OR "barriers" OR "strategy"))

(4) IEEE Xplore: (("All Metadata": "cloud computing" OR "cloud-based") AND ("All Metadata": "biomedical" OR "health research") AND ("All Metadata": "collaboration" OR "consortia") AND ("All Metadata": "barriers" OR "strategy"))

Eligibility criteria and study selection

The inclusion and exclusion criteria were as follows: We included English-language articles and conference papers that explicitly report on the practical implementation or application of cloud computing in biomedical and public health research. We excluded studies concentrating solely on theoretical designs, conceptual architectures, or abstract models without reporting practical applications.

The selection was performed in two phases to ensure both clinical and technical coverage. As previously stated, an initial search of PubMed and Google Scholar was supplemented in February 2026 by a technical expansion into IEEE Xplore and

Web of Science. This comprehensive search strategy yielded a total of 15,604 records (Figure 1). After the removal of duplicates and an initial screening of titles and abstracts, each member of our research team independently reviewed the full texts of the remaining candidate articles.

Eligibility was determined through the assessment of study type, implementation maturity, and direct relevance to the adoption of cloud computing for health research. To prevent the omission of landmark studies, we also performed a backward citation search by reviewing the reference lists of all the eligible papers. This rigorous consensus-based process resulted in a final sample of 68 articles for qualitative synthesis.

Protocol and registration

The study protocol was developed following the PRISMA 2020 statement. Formal registration with International Prospective Register of Systematic Reviews (PROSPERO) was explored, but because the scope of this review centered on computational infrastructure, software applications, and systemic adoption barriers rather than direct health-related outcomes in human participants, it did not meet the eligibility criteria for listing in the registry. All methodological steps were predefined to ensure transparency and minimize bias.

Data extraction, synthesis, and analysis

The data analysis and synthesis were conducted in two distinct but complementary stages: descriptive synthesis using a data extraction

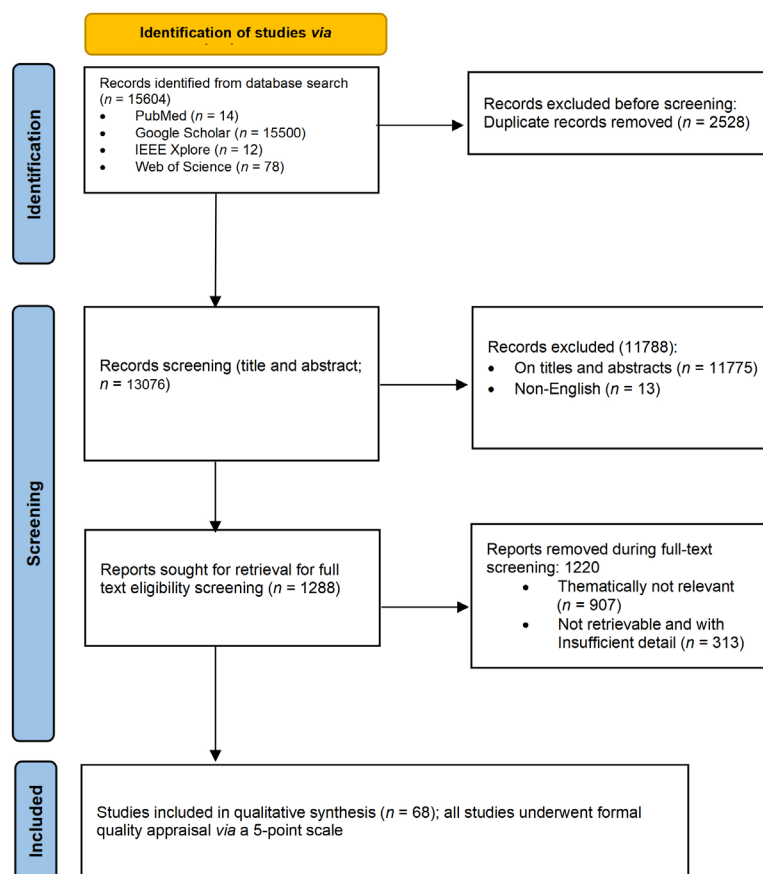


Figure 1. PRISMA flow diagram of the literature search. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

table and conceptual synthesis grounded in a structured analytical technique for constructing a model of systemic barriers.

Descriptive synthesis

Data from the 68 studies were systematically retrieved and summarized using a prepiloted data extraction framework. This process was managed *via* Google Sheets, which facilitated real-time collaboration among the reviewers and ensured standardized data entry across the sample. The extraction table was structured in a way that captured a comprehensive suite of descriptive and technical characteristics for every application. The specific information collected from each study comprised the following. (1) Study identifier: Full citation details and year of publication; (2) Metadata on application: Application name, type (*e.g.*, data sharing platform, analytics pipeline), and primary use/focus (*e.g.*, genomic data analysis, collaboration on medical imaging); (3) Infrastructure and deployment: The host/deployment model used (*e.g.*, public cloud, private cloud, or specific commercial provider); (4) Contextual data: Relevant geographical/funding context, study site, and source of funding (to identify potential institutional or commercial bias); (5) Qualitative insights: Crucial data regarding the principal limitations identified in the primary studies.

This structured extraction process provided the foundational facts regarding application deployment and usage. Capturing these variables, particularly deployment maturity and reported limitations, enabled us to generate the necessary data points with which to conduct the subsequent informatics quality appraisal and prepare the data for subsequent interpretive analysis. The extraction process also guaranteed a comprehensive, granular cataloging of the current cloud landscape for collaborations on biomedical and health research.

Conceptual synthesis

To move beyond descriptive reporting and achieve the review's goal of shedding light on the systemic obstacles to effective cloud-based collaboration on health research, we employed a structured analytical approach rooted in thematic synthesis.^[23] This technique allowed us to transform heterogeneous limitations into an explanatory conceptual framework. The process involved two stages applied to the limitations/gaps section of our extraction table. (1) Axial coding and categorization: The granular, descriptive shortcomings reported for each application (*e.g.*, "susceptible to collusion attacks", "high cost of service", *etc.*) were systematically coded into a concise set of thematic categories (*e.g.*, Security - collusion [S-COLL], Cost to Access [A-COST]). These thematic codes were then grouped under four macro-level pillars: security, access, usability, and functionality; (2) Conceptual model development: The resulting pillars were organized into a conceptual model of systemic barriers, which served as the central analytical tool for visually articulating the complex interdependence between the barriers and explaining why they collectively create a bottleneck to the effective use of cloud computing in collaborative research.

Evaluation of study quality and risk of bias

To address the potential for bias and establish an evidence hierarchy among the reviewed studies, we performed a formal quality appraisal of such research using a customized Internal

Quality Assurance (IQA) framework. Given that traditional clinical risk-of-bias tools, such as Cochrane and Risk Of Bias In Non-randomized Studies - of Interventions (ROBINS-I), are often unsuitable for assessing papers on technical infrastructure and software architecture, the IQA scale was developed to evaluate the maturity, transparency, and validation of cloud-based applications. Each study was independently evaluated by two reviewers on the basis of five weighted domains: (1) Technical documentation - whether a study provided sufficient architectural detail (*e.g.*, VM configurations, Application Programming Interface [API] structures, or containerization) for reproducibility; (2) Security and compliance - whether the study explicitly addressed data sovereignty, security, and regulatory alignment (*e.g.*, the Health Insurance Portability and Accountability Act [HIPAA], the General Data Protection Regulation [GDPR]); (3) Validation status - whether the application was validated using real-world biomedical datasets or limited to a conceptual "sandbox" or synthetic data; (4) Operational maturity - whether the application was a production-ready environment with multiuser support or a localized prototype/tool; and (5) data governance - whether there were clear protocols for data access, identity management, and collaborative auditing.

The studies were assigned a score ranging from 1 (conceptual/Low maturity) to 5 (production-ready/high maturity). This scoring system allowed us to categorize 50 applications into an evidence hierarchy, distinguishing between nascent computational tools and established, high-impact research platforms.

RESULTS

Characteristics of the studies

The 68 studies, which yielded a total of 50 unique applications and archives, are categorized in Table 1 on the basis of research focus, deployment model, and geographical origin. Our analysis transcended simple identification to include the evaluation of the maturity and systemic deficiencies of the implementations documented in the studies, providing a comprehensive view of the current cloud-integrated health landscape.

Quality of the studies

The IQA revealed a remarkably mature evidence landscape. Nearly half of the identified applications (46%, $n = 23$), constituted by production-ready infrastructures, such as Terra, Galaxy, and the UK Biobank, have reached the highest maturity level (Score 5). Only 4% ($n = 2$) were classified as emerging technical tools or conceptual frameworks (Scores 1-2). The remaining applications were distributed across validated implementations and established research tools (Scores 3-4). This distribution indicates that the vast majority of the field (96%) has successfully transitioned from theoretical concepts into validated or production-level research environments.

Temporal and empirical analysis

The evolution of the identified applications (Figure 2) mirrors the broader transformation of healthcare IT over the last 15 years. From 2010 to 2026, the field transitioned from task-specific tools to integrated data commons, such as those serving the cancer and genomics communities, before eventually shifting to the massive,

Table 1**Summary of reviewed cloud-based applications for biomedical and health research**

Study ID	Application	Type	Primary Focus	Deployment	IQA*	Search Source
Afğan <i>et al.</i> ^[24]	Galaxy	Platform	Genomic workflows	IaaS/OS	5	PubMed/IEEE/Scholar
Heath <i>et al.</i> ^[25]	Bionimbus Cloud	Platform	Protected genomics	OpenStack	4	PubMed/IEEE/GS
Dehnavieh <i>et al.</i> ^[26]	DHIS-2	Platform	Health management/EMR	Open source	4	PubMed/WoS/GS
Ananthakrishnan <i>et al.</i> ^[27]	Globus	Platform	Data movement	Platform	5	PubMed/WoS/GS
Dahlquist <i>et al.</i> ^[28]	NIH Biomedical	Platform	FAIR data strategies	Platform	5	PubMed/WoS/GS
Libuit <i>et al.</i> ^[29]	Terra	Platform	Scalable analysis	Cloud native	5	PubMed/WoS/GS
Li <i>et al.</i> ^[30]	Vivli	Platform	Clinical trial data	Azure	4	PubMed/WoS/GS
Mohammed-Rajput <i>et al.</i> ^[31]	OpenMRS	Platform	EMR/EHR management	Open source	4	PubMed/WoS/GS
Fazen <i>et al.</i> ^[32]	AccessMRS	Platform	Quality EMR	Platform	3	PubMed/WoS/GS
Xia <i>et al.</i> ^[33]	Cloud-ECG	Platform	ECG monitoring	Cloud	3	PubMed/IEEE/GS
Skolariki and Avramouli ^[34]	CaTRIP	Platform	Cancer data query	Platform	3	PubMed/WoS/GS
Cerami <i>et al.</i> ^[35]	cBio-Genetics	Platform	Cancer exploration	Platform	4	PubMed/WoS/GS
Ohno-Machado <i>et al.</i> ^[36]	iDASH	Platform	HIPAA-compliant sharing	Cloud	4	PubMed/WoS/GS
Toga <i>et al.</i> ^[37]	GAAIN	Platform	Alzheimer's data	Platform	4	PubMed/WoS/GS
Athey <i>et al.</i> ^[38]	transSMART	Platform	Translational info	Platform	4	PubMed/WoS/GS
Afğan <i>et al.</i> ^[39]	CloudMan	Platform	Infrastructure scaling	AWS	4	PubMed/IEEE/GS
Donkor <i>et al.</i> ^[40]	BLAST	Tool	Genome searching	Web/cloud	5	PubMed/WoS/GS
Sahni <i>et al.</i> ^[41]	CloudBurst	Tool	Parallel read mapping	Public cloud	3	IEEE Xplore/GS
Nguyen <i>et al.</i> ^[42]	CloudAligner	Tool	Sequence mapping	Public cloud	2	IEEE Xplore/GS
Feng <i>et al.</i> ^[43]	Peak Ranger	Tool	ChIPseq peak calling	Public cloud	2	IEEE Xplore/GS
Jourdren <i>et al.</i> ^[44]	Eoulsan	Tool	High-throughput seq	IaaS	3	IEEE Xplore/GS
Zhao <i>et al.</i> ^[45]	Rainbow	Platform	Genomics and clinical	Public cloud	3	IEEE Xplore/GS
Thiruvilwamala <i>et al.</i> ^[46]	CareCloud	Platform	Clinical workflow and practice mgmt	Public cloud	4	PubMed/WoS/GS
Doel <i>et al.</i> ^[47]	GiftCloud	Tool	Secure image archival and sharing	Public cloud	3	WoS/PubMed/GS
Sunyaev <i>et al.</i> ^[7]	Google Health	Platform	PHR management	Public cloud	4	PubMed/WoS/GS
Sunyaev <i>et al.</i> ^[7]	MS HealthVault	Platform	PHR management	Public cloud	4	PubMed/WoS/GS
Hoyt <i>et al.</i> ^[48] Aggarwal <i>et al.</i> ^[49]	IBM Watson	Platform	Health analytics	Private cloud	4	PubMed/WoS/GS
Ko <i>et al.</i> ^[50]	Closha	Platform	WGS analysis	IaaS	3	IEEE Xplore/GS
Agapito <i>et al.</i> ^[51]	Cloud4SNP	Tool	Data mining/SNPs	Cloud	3	IEEE Xplore/GS
Yung <i>et al.</i> ^[52]	Cancer Collab.	Platform	Large-scale cancer	Platform	4	PubMed/WoS/GS
Prior <i>et al.</i> ^[53]	TCIA	Archive	Cancer imaging	DaaS	5	PubMed/WoS/GS
Underwood <i>et al.</i> ^[54]	PCAWG	Archive	Noncoding regions	DaaS	5	PubMed/WoS/GS
Raney <i>et al.</i> ^[55]	Genome Mapper	Archive	Microbio genomes	DaaS	4	PubMed/WoS/GS
Tomczak <i>et al.</i> ^[56]	TCGA	Archive	Cancer variations	DaaS	5	PubMed/WoS/GS
Katz <i>et al.</i> ^[57]	SRA	Archive	Seq data archive	DaaS	5	PubMed/WoS/GS
Lee and Majumder ^[58]	NDA	Archive	Mental health data	DaaS	5	PubMed/WoS/GS
Lonsdale <i>et al.</i> ^[59]	G-Tex	Archive	Tissue/gene var.	DaaS	5	PubMed/WoS/GS
Burgess <i>et al.</i> ^[60]	TOPMed	Archive	Biological processes	DaaS	5	PubMed/WoS/GS
Sathe <i>et al.</i> ^[61]	Enrol-HD	Archive	Proteomics research	DaaS	5	PubMed/WoS/GS
Brown <i>et al.</i> ^[62]	PPMI	Archive	Parkinson biomarkers	DaaS	5	PubMed/WoS/GS
Tryka <i>et al.</i> ^[63]	dbGAP	Archive	Genotype/phenotype	DaaS	5	PubMed/WoS/GS
Toga <i>et al.</i> ^[64]	GAAIN	Archive	Alzheimer's samples	DaaS	5	PubMed/WoS/GS
Mchugh <i>et al.</i> ^[65]	ADNI	Archive	Alzheimer's samples	DaaS	5	PubMed/WoS/GS
Barrett ^[66]	GEO	Archive	Gene expression	DaaS	5	PubMed/WoS/GS
Gaye <i>et al.</i> ^[5]	DataSHIELD	Platform	Multicohort research	Decentralized	4	PubMed/WoS/GS
Conroy <i>et al.</i> ^[67]	Biobank	Platform	Public health research	AWS	5	PubMed/WoS/GS
Lappalainen <i>et al.</i> ^[68]	European Genome-Phenome Archive (EGA)	Archive	Protected genomic-phenomic data	Hybrid cloud	5	PubMed/WoS/GS
Jensen <i>et al.</i> ^[69]	Genomic Data Commons (GDC)	Archive	Genomic Data Commons (NIH Cancer)	Public cloud	5	PubMed/WoS/GS
Zhang <i>et al.</i> ^[70]	International Cancer Genome Consortium (ICGC Portal)	Archive	Int. Cancer Genome Consortium data	Distributed cloud	5	PubMed/WoS/GS
Burgin <i>et al.</i> ^[71]	European Nucleotide Archive (ENA)	Archive	European Nucleotide Archive	Public cloud	5	PubMed/WoS/GS

Characteristics and IQA of cloud-based health applications. This table synthesizes the technical focus, deployment models, and maturity levels of 50 applications. IQA scores are defined on a 5-point maturity scale, where 5 represents production-ready infrastructure, 4 indicates validated implementation, 3 denotes an established research tool, 2 signifies pilot or prototype, and 1 represents a conceptual framework. Abbreviations: IQA, Informatics Quality Appraisal; GS, Google Scholar; WOS, Web Of Science; DHIS, District Health Information Systems; CLOUD-ECG, Cloud-Based Electrocardiogram; GAAIN, Global Alzheimer's Association Interactive Network; BLAST, Basic Local Alignment Search Tool; TCIA, The Cancer Imaging Archive; PCAWG, Pan-Cancer Analysis of Whole Genomes; TCGA, The Cancer Genome Atlas; SRA, Sequence Read Archive; NDA, National Institute Of Mental Health Data Archive; PPMI, Parkinson's Progression Markers Initiative; ADNI, Alzheimer's Disease Neuroimaging Initiative; GEO, Gene Expression Omnibus; IEEE, Institute Of Electrical And Electronics Engineers; WGS, Whole Genome Sequencing; PHR, Personal Health Record; SNPs, Single Nucleotide Polymorphisms.

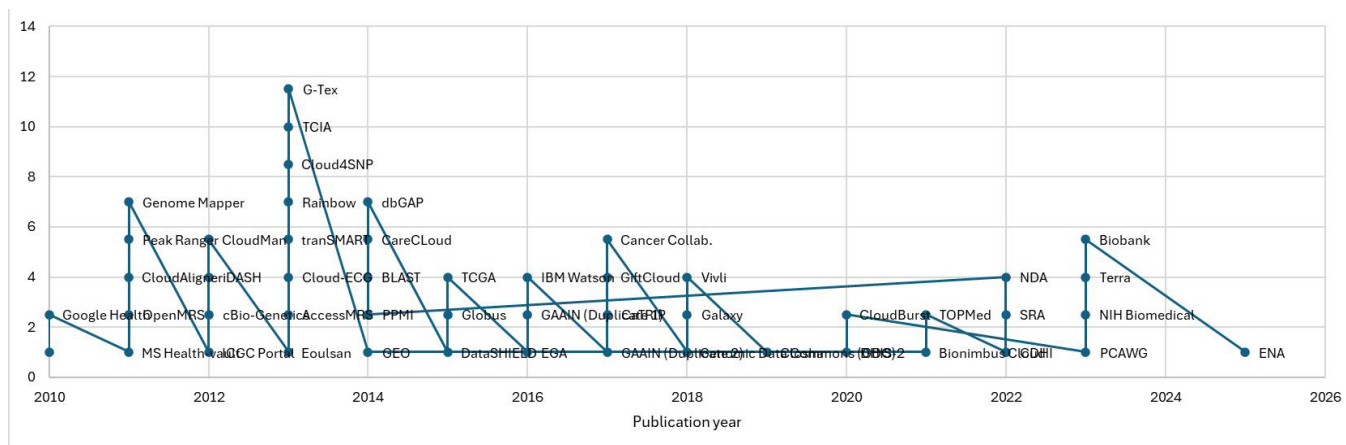


Figure 2. Temporal distribution of cloud applications for health research collaboration (2010-2025). TCGA, The Cancer Genome Atlas; dbGaP, Database of xGenotypes and Phenotypes; EGA, European Genome-phenome Archive; GA4GH, Global Alliance for Genomics and Health; GTEx, Genotype-Tissue Expression Project; GDC, Genomic Data Commons; SRA, Sequence Read Archive; TCIA, The Cancer Imaging Archive; TOPMed, Trans-Omics for Precision Medicine; NDA, National Data Archive; PCAWG, Pan-Cancer Analysis of Whole Genomes; ENA, European Nucleotide Archive; CDHI, Cancer Diagnosis Hub Initiative.

AI-ready ecosystems observed today. This trajectory was driven by the maturation of commercial hyperscalers (AWS, Azure, and Google Cloud) and the escalating data demands of next-generation sequencing (NGS).^[72,73] This trajectory was confirmed by market intelligence for 2025, with the healthcare cloud computing market maintaining a compound annual growth rate of over 17%, projected to exceed \$190 billion by 2032.^[74]

The analysis also revealed a clear maturation of the ecosystem, with a distinct functional viewpoint of cloud application utilization. Despite this reality, however, the studies highlighted that momentum has been highly uneven. For example, the Global Cloud Ecosystem Index,^[75,76] and recent 2025 United Nations Conference on Trade and Development (UNCTAD) reports pointed to a readiness gap.^[77] Whereas leading economies such as Singapore, the USA, and European nations aim for AI-aware and highly protected cloud architectures to enhance security, low- and middle-income countries (LMICs), particularly those in Africa, remain constrained by the current digital baseline. More of this is evidenced by increasing deployment disparity, as determined in the analysis of the studies and cloud applications, posing an inclusion challenge to LMIC researchers. While the identified applications offer powerful collaboration resources, their effective use in LMICs is often affected by slow internet speeds and sparse coverage, high data-egress costs, and regulatory misalignment.^[74] Consequently, issues of data sovereignty and regulatory complexity often erode the trust required for researchers to engage as equal partners in international consortiums.

Geographical distribution of research

The geographical origins of the studies and the applications that they document highlights a clear disparity between innovation leadership and the necessity of global deployment. As illustrated in Figure 3D and the application distribution data table in Appendix 1, the vast majority of initiatives to develop cloud applications for biomedical and health research and large-scale data archive infrastructure originate from North America, primarily the United States. This includes foundational applications such as Cloud-

Burst,^[41,78] and Galaxy^[24] as well as major archives including The Cancer Genome Atlas (TCGA)^[56] and the Bionimbus Cloud.^[25] This concentration suggests that early cloud adoption is driven largely by well-funded, high-volume research consortia in affluent economies.

In contrast, platforms originating from or heavily influenced by European consortia and institutions, such as DHIS-2 and DataSHIELD,^[5,65] demonstrate a strong focus on cross-border collaboration, privacy-preserving techniques, and deployment for global public health management. The widespread adoption of DHIS-2 as a national health system in over 75 LMICs serves as the strongest evidence of cloud applications impacting global health equity.

Applications: tools, platforms, and archives

We categorized the 50 cloud computing applications identified here into three main types: tools, platforms, and archives. This categorization was based on their primary function in the workflow of biomedical and public health research, as summarized in a comprehensive data extraction sheet (see Supplemental Document 1: <http://links.lww.com/DM/A16>). We distinguished cloud applications not only by their underlying stack (Infrastructure as a Service [IaaS]/ platform-as-a-service [PaaS]/ Software as a Service [SaaS]), but by their role in the lifecycle of biomedical and public health research. The analysis uncovered a particularly strong adoption of cloud technology in computationally intensive domains, such as genomics, and data-intensive fields, such as cancer research and translational medicine.

Tools

The analysis of the cloud technologies in Table 1 revealed a special type of applications characterized by their effectiveness in executing a single, well-defined function in the research process, often focusing on high-performance computing tasks. These tools, unlike collaborative ecosystems, are generally designed to ingest raw data for a specific computational task before terminating. As shown in Table 2, a subset of identified applica-

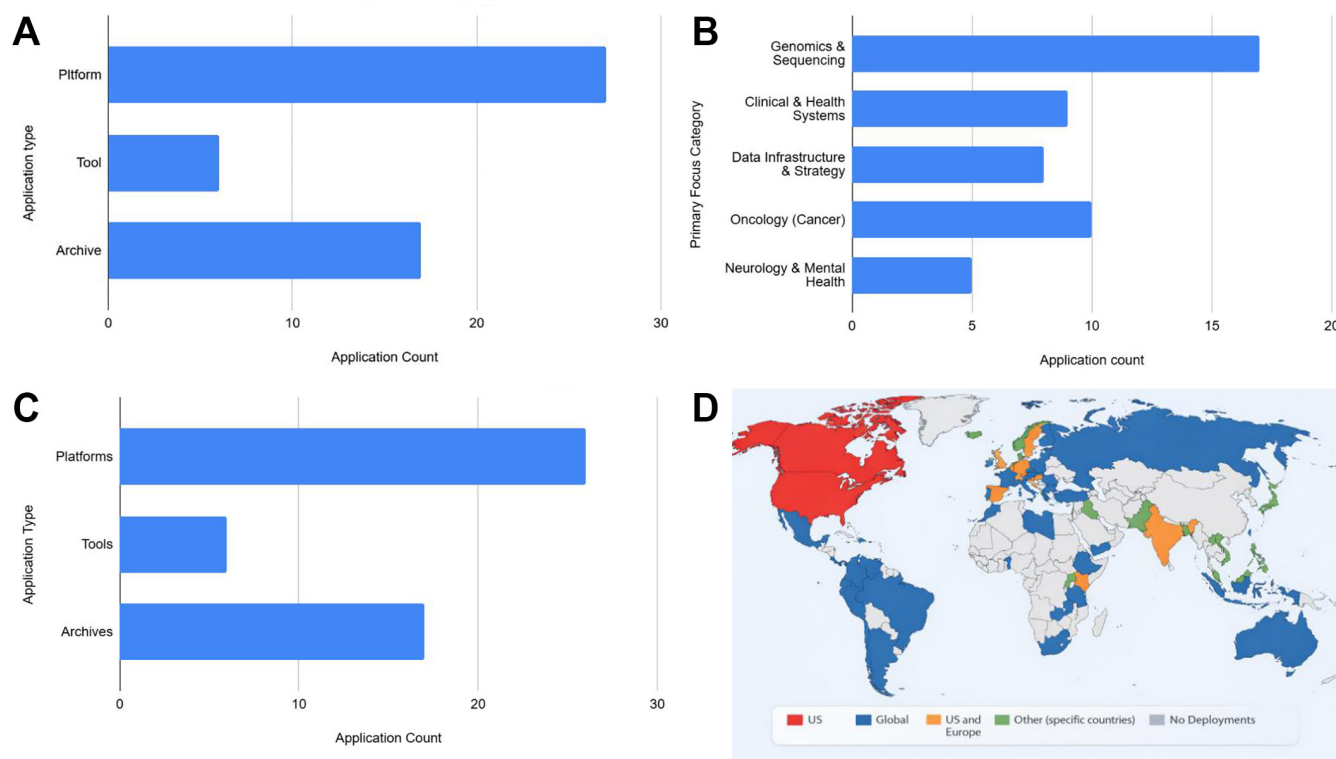


Figure 3. Analyses of the cloud applications based on several criteria. (A) Top-left illustrates the distribution of identified cloud application by type. (B) Top-right shows the distribution of applications by domain focus. (C), (D) Bottom-right and bottom-left display the global spread of development vs. the deployment/use of the applications.

Table 2

Comparative assessment of cloud-based applications for research collaboration - tools

Application	Hosted	Supports data sharing	Flexibility and usability	Confidentiality preserved	Likelihood of collusion
BLAST ^[40]	Public cloud	y	n	n	y
CloudBurst ^[41]	Public cloud	y	n	n	y
CloudAligner ^[42]	Public cloud	y	n	n	y
Peak Ranger ^[43]	Public cloud	y	n	n	y
Eoulsan ^[44]	Public cloud	y	n	n	y
Cloud4SNP ^[51]	Public cloud	y	n	n	y

BLAST, Basic Local Alignment Search Tool.

tions, categorized here as tools (e.g., BLAST^[40]), was evaluated on the grounds of hosting environment/Location, support for data sharing, flexibility, usability, and security (confidentiality and likelihood of collusion). CloudBurst,^[41] CloudAligner,^[42] Peak Ranger,^[43] Eoulsan,^[44] and Cloud4SNP^[51] were evaluated against the critical requirements for secure biomedical collaboration, including hosting environment/Location, support for data sharing, flexibility, usability, and security (confidentiality and likelihood of collusion). These applications demonstrate significant high data-sharing capabilities but lack user-centric flexibility and native confidentiality features. They are often susceptible to the risk of data collusion because they are typically missing integrated privacy-preserving layers. Given that they leverage public cloud scalability through parallelization architectures without encryption-in-use techniques, they remain vulnerable when data fragments are intercepted across distributed cloud nodes.

The majority of the tools are primarily used for genomics

and sequencing owing to massive data generation from NGS. Cloud-based tools such as CloudBurst^[41] and CloudAligner^[42] take advantage of the elastic scalability of the cloud to perform parallel short-read mapping—a critical, high-throughput step in genetic analysis. Similarly, Peak Ranger^[43] applies cloud resources to ChIP-seq data analysis, discovering interactions between proteins and DNA. This reliance on the cloud is essential for managing the complete scale of sequencing data, which would otherwise be computationally prohibitive on local infrastructure. From carrying out general analysis to examining specific health functionality, beyond genomics, cloud-based tools enable general-purpose bioinformatics, as is the case with BLAST,^[40] and specialized data mining, as represented by Cloud4SNP,^[51] for the parallel processing of pharmacogenomics microarray data.

Platforms

Platforms represent integrated, often multitenant environments

that enable a suite of activities, including data management, complex workflow execution, collaboration, and secure access. They are the backbone of interorganizational research consortia. Out of the 50 cloud-based health research applications identified, 26 are platforms (Table 3).

With regard to collaborative bioinformatics, platforms such as Galaxy,^[24,79] CloudMan,^[38,81] and Clossha^[50] are characterized by open-source architectures and web-based environments used for data-intensive biomedical research and large-scale data analysis workflows. These platforms provide user interfaces with which to manage diverse data and analytical tools without requiring advanced computational skills. These functionalities are complemented in Globus^[27] through the advanced data management and secure data transfer critical for multisite collaborations.

In terms of domain specificity, many of the platforms are tailored to specific research areas while prioritizing security and compliance. For example, Bionimbus Cloud^[25] is a protected environment for managing and sharing extensive genomic and phenotypic data, supporting projects such as the resequencing of acute myeloid leukemia genomes. Other domain-specific portals include Global Alzheimer's Association Interactive Network (GAAIN)^[82] for Alzheimer's research and cBio-Genetics for cancer genomics. As for clinical data sharing, iDASH^[36] provides a secure, HIPAA-compliant mechanism for sharing algorithms, while Vivli^[30] focuses on creating a secure analytic platform for de-identified data on clinical trial participants.

With respect to deployment for translation and population health, the platforms address challenges in translating research findings into clinical practice. For example, tranSMART^[38] is a translational informatics platform for collaboration, and DataSHIELD^[5] supports the secure, federated analysis of sensitive data stored across different organizations. In public health, DHIS-2 supports disease surveillance and health management at scale,^[26] while Open Medical Record System (OpenMRS) and an Open Access Medical Record System (AccessMRS) provide frameworks for the establishment of collaborative clinical records.^[31,32] Large-scale population studies are further enabled by platforms such as Enrol-HD^[61] and the UK Biobank.^[67]

Finally, commercial and consumer-facing services such as IBM Watson,^[48] CareCloud,^[45] Google Health,^[6] and Microsoft HealthVault^[6] advance public health management and the creation of personal health records. These are distinguished from individual tools by their ability to maintain integrated data ecosystems over time.

Archives

In biomedical and health research, archives function as dedicated repositories designed to house massive, often publicly available datasets that are essential for comparative analysis, validation research, and meta-analysis. Table 4 lists 18 applications categorized as archives. Although many of these repositories use cloud infrastructure for backend storage, their primary service is providing organized data access rather than a flexible, cloud-computing environment for analysis. Most of the applications currently operate on a "download-to-compute" model, wherein users must transfer data to external platforms or local infrastructure for processing.

In terms of application dominance, a considerable majority of the 18 archives are dedicated to general genomics, clinical health systems, and cancer research. Prominent examples include TCGA,^[56] which holds data associated with over 33 tumor types, and The Cancer Imaging Archive (TCIA),^[53] which focuses on quantitative imaging data. This cancer-specific cluster is further supported by the Genomic Data Commons (GDC)^[70] and the International Cancer Genome Consortium (ICGC) Portal,^[70] which serve as centralized hubs for international oncology data. In the domain of general genomics, the Sequence Read Archive (SRA)^[57] is a primary international archival resource for NGS data, complemented by the European Nucleotide Archive^[71] and the European Genome-phenome Archive^[68] for protected genomic-phenomic datasets.

With respect to specificity, beyond use in high-volume sequencing, archives have been established for specific health conditions, demonstrating the targeted application of cloud storage. These include the National Institute of Mental Health (NIMH) Data Archive,^[58] used to store research data on mental health, Parkinson's Progression Marker Initiative (PPMI)^[70] for Parkinson's disease biomarkers, and GAAIN^[82] for aggregating data on Alzheimer's disease, including Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and genetics. Other specialized portals, such as the Alzheimer's Disease Neuroimaging Initiative ADNI Workbench^[65] and G-Tex,^[59] provide tissue-specific expression data and longitudinal insights on Alzheimer's disease. This trend indicates that while the cloud can host general data (*e.g.*, Gene Expression Omnibus (GEO) for functional genomics), its implementation often hinges on the needs of considerably specialized, high-value research cohorts. While the 18 archives provide the necessary motivation for research, they are characterized by relatively reliable security and confidentiality while frequently being confronted with constraints in flexibility because data are often siloed within controlled-access environments.

The analysis of the applications pointed to clear trends that could be further synthesized through visualization. Figure 3A (distribution by type) illustrates the structural focus of cloud solutions, showing the balance between the tools, platforms, and archives. Figure 3B (distribution by domain) highlights functional focus, confirming the dominance of genomic and cancer research in cloud adoption. Most crucially, the geographical data points, summarized in Figures 3C and 3D, demonstrated that while development often originates from high-income countries (*e.g.*, the US for Galaxy and TCGA, Norway for DHIS-2), deployment and impact are considerably global, especially for platforms such as DHIS-2, which is now a national health system in over 75% of LMICs.

Synthesis of application gaps

Following the established methodology, the key limitations and gaps across the identified cloud computing applications were synthesized into thematic codes (Table 5) as well as the code keys (Table 6). This process revealed a tremendous concentration of recurrent challenges that were systematically grouped under four main conceptual pillars: security (S), access (A), usability (U), and functionality (F). The subsequent sections present the analysis of how these pillars systemically restrict the flow of effective collaboration, as visualized in the conceptual model.

Table 3**Comparative assessment of cloud-based applications for research collaboration - platforms**

Application	Supports secure data sharing	Flexibility and usability	Confidentiality preservation	Access status	Likelihood of collusion
Galaxy ^[24,79]	High	Moderate	High	Open source	Moderate
Bionimbus Cloud ^[25]	High	Moderate	High	Open source	Moderate
DHIS-2 ^[26]	High	Low	High	Proprietary/Licensed	High
Globus ^[27]	High	Low	High	Proprietary/Licensed	High
NIH Biomedical ^[80]	High	High	High	Proprietary/Licensed	Low
Terra ^[29]	High	High	High	Open-source model	Low
Vivli ^[30]	High	Low	High	Proprietary/Licensed	High
OpenMRS ^[31]	High	Low	High	Open source	Moderate
AccessMRS ^[32]	High	Low	High	Open source	Moderate
Cloud-ECG ^[33]	High	Moderate	High	Open source	Moderate
CaTRIP ^[34]	High	Low	High	Open source	Moderate
cBio-Genetics ^[35]	High	Moderate	High	Open source	Moderate
iDASH ^[36]	High	High	High	Open source	Moderate
tranSMART ^[37]	High	Moderate	High	Open source	Moderate
CloudMan ^[40]	High	Moderate	High	Open source	Moderate
Rainbow ^[45]	High	Moderate	High	Open source	Moderate
CareCloud ^[46]	High	Low	High	Proprietary/Licensed	Low
Google Health ^[7]	High	Low	High	Proprietary/Licensed	Low
MS HealthVault ^[7]	High	Low	High	Proprietary/Licensed	Low
IBM Watson Health ^[49]	High	Low	High	Proprietary/Licensed	Low
Closha ^[50]	High	Moderate	High	Proprietary/Licensed	Moderate
Cancer Collab. ^[80]	High	Moderate	High	Open source	Moderate
Enrol-HD ^[61]	High	Low	High	Controlled	Low
DataSHIELD ^[5]	High	Moderate	High	Open source	Low
UK Biobank (RAP) ^[67]	High	High	High	Controlled	Low
GiftCloud ^[47]	High	High	Moderate	Controlled	Moderate

This table summarizes the 26 platforms used for biomedical and public health research collaborations. Unlike the task-specific tools, these platforms were evaluated based on their effectiveness in maintaining a collaborative state. DHIS-2, District Health Information System.

Table 4**Comprehensive assessment of cloud-based health applications - archives (*n* = 18)**

Application	Category	Primary focus	Disease specific	Data sharing	Security	Open access
TCIA ^[53]	Archive	Quantitative cancer imaging	Yes	High	High	No
PCAWG ^[54]	Archive	Pan-cancer noncoding genome	Yes	High	High	No
Genome Mapper-ENCODE ^[55]	Archive	Microbial& human (1000GP/ENCODE)	No	High	High	No
TCGA ^[56]	Archive	Cancer genomic variations	Yes	High	High	No
SRA ^[57]	Archive	Raw sequence data (BioData Catalyst)	No	High	High	No
NDA (NIMH) ^[57]	Archive	Mental health research data	Yes	High	High	No
G-Tex ^[59]	Archive	Tissue-specific gene expression	No	High	High	No
TOPMed ^[60]	Archive	Heart, lung, blood precision medicine	Yes	High	High	No
PPMI ^[62]	Archive	Parkinson's biomarkers	Yes	High	High	No
dbGAP ^[63]	Archive	Phenotype& genotype interaction	No	High	High	No
AD Workbench ^[65]	Archive	Alzheimer's data portal& analytics	Yes	High	High	No
GEO ^[66]	Archive	Functional genomics (public)	No	High	High	Yes
EGA ^[68]	Archive	Protected genomic-phenomic data	No	High	High	No
GDC ^[69]	Archive	Genomic Data Commons (NIH Cancer)	Yes	High	High	No
ICGC Portal ^[70]	Archive	Int. Cancer Genome Consortium data	Yes	High	High	No
ENA ^[71]	Archive	European Nucleotide Archive	No	High	High	No
GAAIN ^[64]	Archive	Alzheimer's disease data aggregator	No	High	High	No
Enrol-HD	Archive	Proteomics research	No	High	High	No

Abbreviations: TCIA, The Cancer Imaging Archive; 1000GP, 1000 Genomes Project; AD, Alzheimer's Disease; dbGAP, Database of Genotypes and Phenotypes; EGA, European Genome-phenome Archive; ENA, European Nucleotide Archive; ENCODE, Encyclopedia of DNA Elements; GAAIN, Global Alzheimer's Association Interactive Network; GDC, Genomic Data Commons; GEO, Gene Expression Omnibus; GTEx, Genotype-Tissue Expression; HD, Huntington's Disease; ICGC, International Cancer Genome Consortium; NDA, National Institute of Mental Health Data Archive; NIH, National Institutes of Health; NIMH, National Institute of Mental Health; PCAWG, Pan-Cancer Analysis of Whole Genomes; PPMI, Parkinson's Progression Markers Initiative; SRA, Sequence Read Archive; TCGA, The Cancer Genome Atlas; TOPMed, Trans-Omics for Precision Medicine.

Table 5

Functional classification, deployment architectures, and systemic thematic barriers of selected cloud-integrated health informatics applications and data repositories

Selected applications	Application type	Primary use/focus	Host/deployment model	Key systemic barriers (coded)
UK Biobank	Platform	Population health& genomics	AWS/DNAexus	A-REST, A-COST, U-FLEX, S-REGL
Galaxy	Platform	Genomic workflows	IaaS/Open source	S-COLL, S-REGL, U-INTER, F-REPRO
Bionimbus Cloud	Platform	Protected genomic data	IaaS/OpenStack	S-REGL, U-INTER
DHIS-2	Platform	Public health& EMRs	Platform/OS	U-FLEX, S-COLL, A-CLOSED, F-LACK
Globus	Platform	Data movement/management	Platform	U-FLEX, S-COLL, A-CLOSED
NIH Biomedical Cloud	Platform	Fair data implementation	Platform/strategy	S-COLL, A-COST, A-REST, U-FLEX
Terra	Platform	Scalable open analysis	Cloud native	S-COLL, A-COST, U-FLEX, S-REGL
Vivli	Platform	Clinical trial data sharing	MS Azure	S-COLL, A-REST, A-COST, U-INTER
OpenMRS	Platform	EHR/EMR management	Platform/OS	U-FLEX, S-COLL, F-LACK
AccessMRS	Platform	EMR management (LMICs)	Platform	S-COLL, F-ADOPT, U-INTER
Cloud-ECG	Platform	Real-time monitoring	Cloud-based	F-ADOPT, S-COLL, S-TRUST
iDASH/GAIAIN	Platform	Disease-specific (AD)	Cloud-based	F-CLOSE, S-COLL, A-COST, U-FLEX
tranSMART	Platform	Translational informatics	Platform	F-CLOSE, S-COLL, U-FLEX, A-COST
CloudMan	Platform	Infra distribution	AWS (public cloud)	S-COLL, U-INTER, S-REGL
BLAST	Tool	Genome searching	Cloud/web	S-COLL, U-INTER, F-REPRO
CloudBurst	Tool	Sequence mapping	Public cloud	U-FLEX, S-COLL
CloudAligner	Tool	Sequence alignment	Public cloud	S-COLL, U-INTER, U-FLEX
Peak Ranger/Eoulsan	Tool	Sequencing tasks	Public cloud	S-COLL, A-COST, U-INTER
SoapSNP/CloudSNP	Tool	Sequence analysis	Public cloud	S-COLL, U-INTER, U-FLEX, F-REPRO
PHR (Google/MS/IBM)	Platform	Personal health records	Public/private cloud	S-TRUST, F-ADOPT, U-FLEX
Crossbow	Tool	WGS pipeline	IaaS environment	S-COLL, U-INTER
Cloud4SNP	Tool	Data mining/SNPs	Cloud-based	A-COST, S-TRUST
Cancer Collaboratory	Platform	Large-scale cancer research	Platform/archive	U-FLEX, S-COLL, A-CLOSED, A-COST
TCIA/dbGAP	Archive	Imaging& GWAS data	Cloud storage (DaaS)	A-REST, U-INTER, S-COLL, U-FLEX
PCAWG/TCGA	Archive	Cancer genomics	Cloud storage (DaaS)	U-FLEX, S-COLL, F-CLOSE, A-COST
SRA/GEO	Archive	Genomic/expression data	Cloud storage (DaaS)	S-COLL, A-COST, U-INTER, A-REST
NDA/G-TeX	Archive	Mental health/tissues	Cloud storage (DaaS)	A-REST, U-INTER, S-COLL, A-COST
PPMI	Archive	Parkinson's biomarkers	Cloud storage (DaaS)	A-REST, A-COST, F-REPRO, A-DATA
ADNI Archive	Archive	Alzheimer's imaging	Cloud storage (DaaS)	U-INTER, F-REPRO, S-TRUST
DataSHIELD	Platform	Multicohort research	Decentralized	U-FLEX, A-COST, U-INTER

AWS, Amazon Web Services; N.B.: While the analyzed applications solve certain security issues (S-REGL), they amplify access restrictions (A-REST) and technical complexity for researchers moving local pipelines to the cloud (U-FLEX). Included in the table: AD: Alzheimer's Disease; NIH Biomedical Cloud, OpenMRS, Open Medical Record System; AccessMRS, Access Medical Record System; CLOUD-ECG, Cloud-based Electrocardiogram; iDASH, Integrating Data for Analysis, Anonymization, and Sharing; GAIAIN, Global Alzheimer's Association Interactive Network; tranSMART, Translational Medicine Analysis and Reporting Platform; CloudMan, Cloud Manager; BLAST, Basic Local Alignment Search Tool; SoapSNP, Short Oligonucleotide Analysis Package Single Nucleotide Polymorphism; CloudSNP, Cloud-based Single Nucleotide Polymorphism; PHR, Personal Health Record; Cloud4SNP, Cloud for Single Nucleotide Polymorphism; Cancer Collaboratory, Cancer Collaboratory; TCIA, The Cancer Imaging Archive; dbGAP, Database of Genotypes and Phenotypes; PCAWG, Pan-Cancer Analysis of Whole Genomes; TCGA, The Cancer Genome Atlas; SRA, Sequence Read Archive; GEO, Gene Expression Omnibus; NDA, National Institute of Mental Health Data Archive; GTeX, Genotype-Tissue Expression; PPMI, Parkinson's Progression Markers Initiative; ADNI Archive, Alzheimer's Disease Neuroimaging Initiative Archive; AWS, Amazon Web Services.

DISCUSSION

The systematic review uncovered progress in automation and digital transformation in biomedical and public health research - evidence of demand for efficiency in health research collaborations. This transformation has been characterized by a fundamental shift from localized, high-performance computing toward scalable, network-accessible cloud resources. The findings, particularly the dominance of the PaaS and data-as-a-service (DaaS) models (Table 1), provide a critical context for discussing the current state of adoption, limitations, and future opportunities.

The evolving landscape of cloud adoption

With the diversity of applications presented in Section 3, ranging from highly specialized tools (*e.g.*, CloudBurst, Peak Ranger) to comprehensive platforms (*e.g.*, Galaxy, Terra) and

massive data archives (*e.g.*, TCGA, SRA), the review demonstrated the extensive functional adoption of the cloud paradigm. The review also indicated that the cloud has been primarily adopted as a data repository and analysis environment (PaaS/DaaS) rather than a purely infrastructural solution (IaaS). Researchers require systems that abstract underlying computational complexity, ensuring easy-to-use analysis workflows (*e.g.*, Galaxy, tranSMART) and secure data access (*e.g.*, NIH Biomedical Cloud, Vivli). This trend has been driven by the enormous size and complexity of modern omics data, for which petabyte-level storage or transfer is often impractical due to cost. With cloud technology, the computational workload is passed onto data archives.

Furthermore, adoption has been markedly uneven across geographical and funding landscapes. A significant majority of the robust, large-scale platforms (Terra, Bionimbus, TCGA Collaboratory) are initiatives rooted in or heavily funded by

Table 6**Conceptualization code keys**

Pillar	Code	Theme-relevant limitation	Definition
Security& trust (S)	S-COLL	Collusion/Re-ID Risk	Vulnerability to linkage or correlation attacks, wherein data can be re-identified through multiple sources
	S-REGL	Regulatory friction	Overhead caused by rigid HIPAA, FISMA, or GDPR configurations and the burden of legal agreements (SLA/BAA)
	S-TRUST	Privacy/trust deficit	Lack of patient or provider trust in third-party servers, fear of data breaches in commercial cloud environments
Access& cost (A)	A-COST	Economic barriers	Financial costs including high computational/storage fees, data egress charges, and personnel costs for data curation
	A-REST	Restrictive governance	Barriers created by institutional policies, IP restrictions, and lengthy approval delays (DUA/IRP)
	A-CLOSED	Non-open access	Systems not open access or restricted to specific institutional communities/funders
	A-DATA	Data use constraints	Legal or ethical limits on how data can be used (<i>e.g.</i> , sensitive historical data or consent restrictions)
Usability& interoperability (U)	U-FLEX	Technical flexibility	Lack of architectural flexibility; rigid configurations, WDL complexity, or CLI-driven barriers that alienate users
	U-INTER	Interoperability gap	Failure to integrate with external systems due to lack of FHIR standards, metadata fragmentation, or multicloud barriers
Functionality& adoption (F)	F-LACK	Lacking research tools	A design gap wherein clinical systems (EHR/EMR) lack the advanced analytical functionality needed for research
	F-CLOSE	Closed scope	Functionality locked to a specific disease (<i>e.g.</i> , cancer only) or a single data modality
	F-REPRO	Reproducibility gaps	Technical failures in versioning, alignment errors, or metadata gaps that prevent study replication
	F-ADOPT	Organizational inertia	Resistance to adoption due to "extra work" perception, skill gaps, or intermittent connectivity in LMIC settings

HIPAA, Health Insurance Portability and Accountability Act; FISMA, Federal Information Security Management Act; GDPR, General Data Protection Regulation; SLA, Service Level Agreement; BAA, Business Associate Agreement; IP, Intellectual Property; DUA, Data Use Agreement; IRP, Institutional Review Process; EHR, Electronic Health Record; and EMR, Electronic Medical Record.

US commercial entities and government research bodies such as the NIH and the National Cancer Institute NCI. Conversely, while global applications such as the DHIS-2 and OpenMRS exist, they often revolve around essential public health records (EHR/EMR) instead of extensive genomic analysis. This finding calls attention to a potential availability and accessibility gap in developing nations, where public cloud and bandwidth shortcomings remain rampant.

Critical limitations and challenges to cloud adoption

Conceptual model of barriers

Figure 4 presents our conceptual model of systemic barriers, derived from the analysis of the coded applications.

Conceptualization of systemic barriers to effective cloud-based collaboration

The findings on the 50 cloud-based applications, synthesized in Table 5, revealed that the challenges hindering effective cloud adoption in collaborations on biomedical and health research are not isolated technical failures but constitute a systemic barrier model. This model is built upon the four highly interdependent pillars of security, access, usability, and functionality, which collectively restrict the flow of robust, secure, and scalable collaboration activities. To interpret well the above model and the explanations that follow, we provide the codes and their interpretations in Table 6.

Security and trust barriers (S)

This theme represents the foundational obstacles to data sharing, whose resolution is nonnegotiable. Collaborative studies rely entirely on the ability to share sensitive data securely, yet the current cloud landscape is defined by substantial security and

trust deficits. The prevalence of the S-COLL code (collusion risk) and linkage risks across many of the analytical applications (*e.g.*, tools (BLAST, CloudBurst) and platforms (Galaxy, NIH Cloud) indicates a systemic failure to implement confidentiality-preserving computation at the core level, especially during the analysis phase. Furthermore, the S-TRUST code (trust/privacy failure) punctuates a socioethical barrier, as evidenced by the historical adoption failures of major PHR platforms, such as Google Health and IBM Watson, and the tremendous fear surrounding protected health information (PHI) and compliance with the GDPR. An important consideration, however, is that the need for complex service level agreements (SLAs) and the regulatory burden are substantial deterrents to cloud migration, reinforcing the reluctance to share or store sensitive genomic or clinical data on third-party infrastructure.

Consistent with the literature outside the reviewed studies such as^[83] and others, the evaluated research identified concerns surrounding the security, privacy, and PHI- and genomic data-related regulatory compliance as the primary hindrance to cloud adoption. The findings also point to the stringency of requirements under regulations such as the HIPAA, Federal Information Security Management Act (FISMA), and General Data Protection Regulation (GDPR), which create a considerable compliance overhead that often stifles innovation. This burden is laid out as follows: (1) In platforms such as Bionimbus and Galaxy, the data highlight rigid configuration requirements as a significant hurdle. These security standards, while necessary for data protection, frequently conflict with the technical flexibility required for rapid, multi-institutional data sharing. This rigidity transforms the cloud from a dynamic collaborative space into a constrained environment where administrative hurdles outpace scientific discovery. (2) This barrier is driven primarily by the persistent institutional fear of data breaches or the noncompliance penalties associated

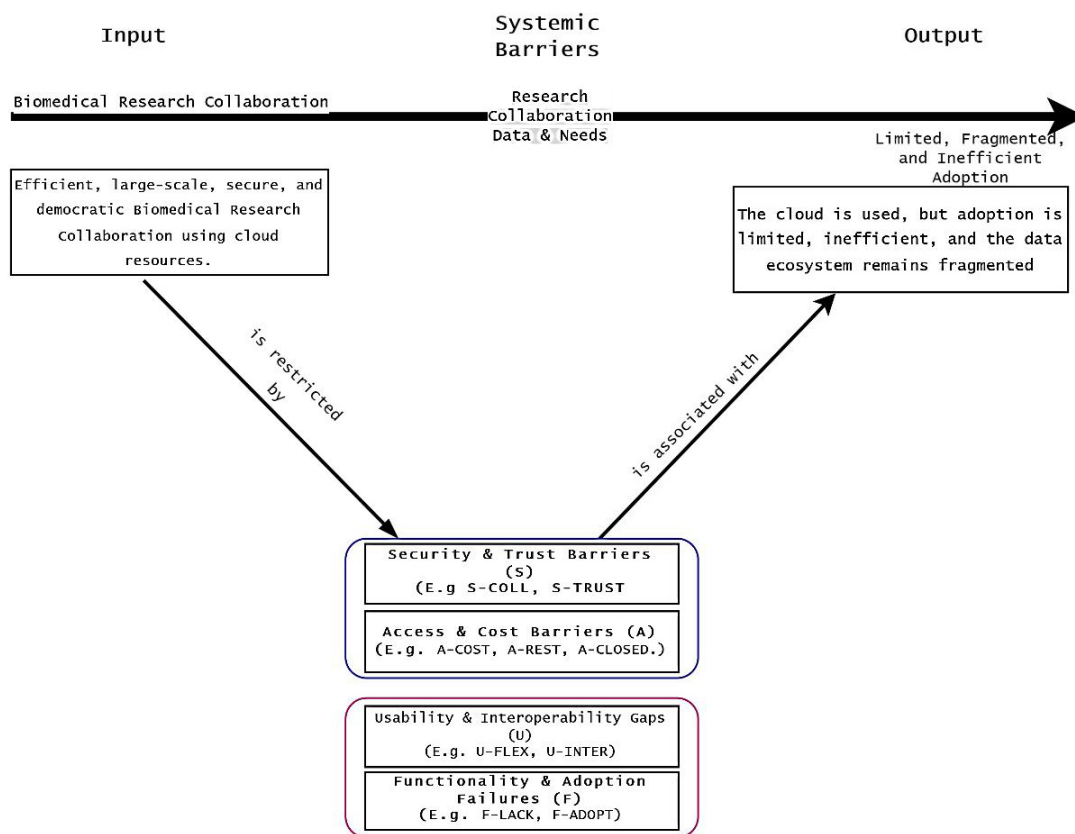


Figure 4. Conceptual model of systemic barriers.

with residing on third-party servers. This worry is not merely theoretical; it is evidenced by the emergence of specialized applications such as iDASH and the NIH Biomedical Cloud, which were engineered specifically for navigation or operation within these complex, compliance frameworks. (3) This landscape necessitates a compliance-first strategy, which compounds planning complexity, budget requirements, and the need for exhaustive legal documentation, such as Service level Agreement (SLAs) and business associate agreements (BAAs). Consequently, the legal and financial burdens of maintaining compliance substantially deter cloud migration, particularly for smaller consortia that lack the legal infrastructure to manage such high overhead.

Access and cost barriers (A)

The second pillar describes the economic and policy factors that constrain participation and shape the cloud ecosystem. The prevalence of the A-COST code confirms that the pay-as-you-go model is often prohibitively expensive for large-scale or long-term consortia, with egress fees and exorbitant storage expenses discouraging the multicloud strategies and collaborations characterizing projects such as Terra, Vivli, and the NIH Biomedical Cloud, among others. More critically, the consistent appearance of A-REST (restrictive governance) and A-CLOSED (restricted access) in major archives (e.g., dbGaP, NIH Biomedical Cloud) reflects a policy barrier. Access is frequently granted only to senior-level PIs following a lengthy data access committee review or evaluation by an independent review panel (IRP) for a data use agreement, effectively excluding junior researchers and resource-limited institutions while also causing data fragmentation instead of fostering real collaboration.

Although the cloud promises cost savings *via* the pay-as-you-go model, this frequently proves true only for burst computing or small-scale projects. For expansive biomedical consortia, the primary cost barrier is data egress—the fee charged to move data out of the cloud. This expense discourages collaboration, data sharing, and multicloud strategies, thus engendering data lock-in. Cost models for sizable archives (DaaS) can be complex and unpredictable, making long-term financial planning difficult for research collaboration grants.

Usability and interoperability gaps (U)

This pillar centers on the technical friction that prevents tools and platforms from being used together efficiently. The prevalence of the U-FLEX code, especially with mechanisms such as the complexity of workflow description language (especially over platforms such as Terra) and barriers related to the command line interface (e.g., CloudAligner) indicates a fundamental absence of flexibility and usability, with tools optimized for specific environments but failing to integrate with others. This situation is exacerbated by vendor lock-in, wherein migration between cloud providers (AWS, Azure) is technically challenging due to differing APIs and proprietary backends. Furthermore, the presence of the U-INTER code highlights issues beyond simple integration, instead encompassing the scalability gap for general tools (e.g., BLAST) and the disparity in efficiency when implementing sophisticated privacy-preserving techniques. These currently require excessive computational resources (compute time and memory) to be practical for massive genomic datasets.

The identified applications use varied proprietary cloud backends (AWS, Azure, Google Cloud). Migrating complex,

highly optimized tools (*e.g.*, Crossbow or CloudAligner) from one cloud platform to another necessitates technical expertise because of variations in APIs and infrastructure configurations. This challenge motivates vendor lock-in, forcing consortia to remain with a single provider and potentially stifling competitive pricing and innovation. In addition, legacy health systems such as TCIA and AccessMRS are often missing the interoperability standards (*e.g.*, Fast Healthcare Interoperability Resources (FHIR) and meta data fragmentation) needed to communicate with modern cloud EHR platforms (Google Health, Microsoft HealthVault). Without standard data exchange protocols, the cloud therefore remains a collection of isolated data silos rather than a unified research ecosystem.

Functionality and adoption failures (F)

The final pillar addresses the divide between technical availability and real-world utility in existing workflows for research collaboration. The cloud applications surveyed typically fail not because they are technically unsound but because they are ill-suited to a user's practical needs. The F-LACK code (lacking research tools), dominant in EMR systems (DHIS-2, OpenMRS), confirms a design barrier: These platforms are optimized for clinical administration but lack the advanced analytical and cohort-building features necessary for research. Similarly, F-CLOSE (closed scope) restricts data utility to a single disease or modality (*e.g.*, CaTRIP). Finally, F-ADOPT (adoption failure), as seen in Google Health and IBM Watson, emphasizes a workflow obstacle, in which even secure platforms struggle to gain traction because of organizational inertia, and a skill gap, which prevents seamless and efficient integration into clinical or research routines.

Other critical deficiencies: skill gap and organizational inertia

Notwithstanding the technical benefits of cloud technologies, the path to widespread, seamless adoption is also challenged by the skills gap and organization inertia that challenge a user's decision to migrate. Cloud adoption is not a purely technical exercise but requires socio-organizational transformation. The tools are often sophisticated, calling for specialized skill sets in cloud architecture, distributed systems, and data engineering—competencies that are conventionally rare in biomedical research teams. Organizational inertia, including resistance to cultural change and the need to decommission legacy systems, further slows down migration.

Opportunities for future cloud biomedical research

The unique characteristics of cloud computing, such as scalability, accessibility, and elasticity, present four major opportunities for the future of biomedical and global health research. Foremost among these is the inclusion of Artificial Intelligence/Machine Learning (AI/ML) and predictive analytics in health scholarship, as the cloud is the only viable infrastructure for training and deploying machine learning models on petabyte-scale data archives. By co-locating DaaS offerings such as TCGA with PaaS environments such as Terra, researchers can efficiently leverage AI for advanced application.

These capabilities directly facilitate precise diagnostics and prediction of health outcomes. For instance, using cloud-native tools to analyze pathology images enables clinicians to significantly improve diagnostic precision, and the ability to forecast

patient risk on the basis of real-time data or integrated genomic profiles allows for a more proactive and personalized approach to medicine in 2026 and beyond.

Cloud infrastructure also serves as an important driver of global health equity and the expansion of telemedicine. Applications such as DHIS-2 and cloud-based EHRs such as OpenMRS demonstrate the considerable opportunity to elevate healthcare delivery in LMICs. By eliminating the need for expensive, localized IT infrastructure, the cloud clears the way for the cost-effective deployment of real-time disease surveillance and remote clinical care, fostering an environment conducive to true global health collaboration.

Finally, the cloud drastically accelerates drug discovery by hosting massive, pre-analyzed datasets alongside powerful computational tools. Platforms that house resources such as dbGAP, G-Tex, and Vivli allow researchers to perform complex simulations and analyze clinical trial data *via* environments. This integration enables dispersed teams to collaborate with unprecedented efficiency, effectively shortening the development timeline.

Recommendations

On the grounds of the synthesis of the reviewed applications, limitations, and opportunities, several key recommendations were formulated for research consortia and individuals intending to transition their workloads and datasets to the cloud in 2026. Primarily, organizations should adopt a compliance-first and phased migration strategy, acknowledging that successful cloud adoption begins with regulatory due diligence rather than simply technical planning. This involves comprehensive compliance mapping to align with standards such as HIPAA and GDPR, followed by a phased rollout using noncritical, de-identified datasets to refine security and workflows while minimizing operational downtime.

To manage economic risks, stakeholders must strategize cost management by minimizing data egress through a shift to compute-to-data operational models. By prioritizing PaaS or DaaS solutions such as Terra or Galaxy, researchers can run analyses directly on co-located archives, including TCGA or SRA. This approach, implemented in conjunction with the proactive use of cloud provider cost management tools to monitor budgets and restrict unnecessary data movement, eliminates egress fees and advances sustainable research.

Another essential measure is to invest in cloud literacy and broader stakeholder engagement to overcome organizational skill gaps. This requires dedicated funding for the targeted training of research staff in cloud-native architectures and best practices for cost management. Successfully implementing technical upgrades likewise depends on engaging clinical and administrative staff early in the planning process. Without their direct involvement and training on new workflows, the practical adoption of systems such as OpenMRS, and AccessMRS is likely to be met with significant friction.

Finally, consortia must prioritize data cleansing and interoperability to prevent legacy system-associated complexities from spilling over to the new environment. Premigration efforts should encompass rigorous data assessment and validation to

remove duplicates and correct errors before extraction is carried out. Accurate data mapping is equally vital, ensuring that all metadata including timestamps, diagnostic codes, and unique identifiers are precisely aligned with the schema of the destination cloud platform. This maintains long-term data integrity and system interoperability.

CONCLUSION

Limitations

While our systematic review followed a rigorous methodology adhering to PRISMA 2020 guidelines, we recognize that the specific boundaries of any search strategy introduce inherent limitations. A primary consideration is the tension between capturing broad organizational governance and deeply structured technical architecture. Our strategy was intentionally designed to prioritize the literature on platforms and governance, focusing on collaborative strategies and consortia models. While this approach is highly effective for documenting adoption patterns in formal institutional structures such as that typifying the NIH, it initially risked the under sampling of emerging technical frameworks, such as specific containerization or serverless applications, often documented primarily in technical proceedings.

To address this shortcoming, we expanded our search, specifically targeting high-yield technical databases, but a critical drawback remains: Next-generation interoperability standards or novel efficiency mechanisms (*e.g.*, those addressing the U-INTER gap) may still be underrepresented if explorations remain confined to the theoretical literature on computer science and have not been piloted in the biomedical domain. Ultimately, the choice to maintain exceptional specificity for collaborative health research represents a calculated trade-off. By focusing on implemented platforms in lieu of purely theoretical models, we ensured that our synthesis reflects real-world systemic barriers, providing an evidence-based map of the adoption landscape as it exists in practice today.

Final remarks

Our systematic synthesis of 68 studies revealed a remarkably mature cloud ecosystem, with 96% of the 50 identified applications having transitioned from conceptual frameworks into validated, production-ready research environments. High-performance tools such as CloudBurst, large-scale archives like the GDC, and integrated platforms including Terra and Galaxy demonstrate that cloud computing is now central to biomedical discovery. This transition has been powered by a market projected to exceed \$190 billion by 2032, driven by the inescapable necessity of managing the exponential growth of NGS and EHR data through AI-ready architectures.

Nevertheless, our analysis also uncovered a significant maturity-accessibility paradox. While the technical quality of these applications is exceptional, as evidenced by 46% of the systems achieving the highest IQA score, their global reach remains profoundly uneven. The geographical concentration of innovation in North America has created a "walled garden" effect, wherein high-value platforms are often constrained by high data egress costs (A-COST), restrictive accessibility (A-REST), and regulatory misalignments (S-REGL). In contrast,

the success of European-led initiatives such as DHIS-2 confirms that cloud applications can drive global health equity when designed with cross-border interoperability and local resource constraints in mind.

Ultimately, the future of cloud-integrated health research hinges on closing the discrepancy between trust and readiness that we identified in our synthesis. Moving forward, the priority must shift from building purely powerful systems to building inclusive ones. To prevent a deepening digital divide, the next generation of cloud development must transcend the "download-to-compute" model and move toward federated, privacy-preserving architectures that address the U-INTER (interoperability) and S-COLL (collusion) risks determined in our thematic analysis. Only by intentionally designing for global accessibility and data sovereignty can we ensure that the cloud fulfills its promise as a truly democratic engine for public health and biomedical collaboration.

Acknowledgements

None.

Author contributions

Mugonza R: conceptualization, methodology, validation, formal analysis. Mugonza R, Ssembatya R, Basaza-Ejiri A: investigation. Mugonza R: writing—original draft preparation, writing—review and editing, visualization. Ssembatya R, Basaza-Ejiri A: supervision. All authors have read and agreed to the published version of the manuscript.

Source of funding

This research received no external funding.

Informed consent

Not applicable.

Ethical approval

Not applicable.

Conflict of interest

The authors have no conflicts of interest to declare.

Use of large language models, AI and machine learning tools

During the final proof reading, some graphs were made clearer rather than blurred images using ChartGenAI. After the generation, using data provided to them, the authors reviewed the final plots and fully take responsibility for the published content.

Data availability statement

No additional data.

REFERENCES

1. Yang X, Wallom D, Waddington S, *et al.* Cloud computing in e-Science: Research challenges and opportunities. *J Supercomput.*

- 2014;70(1):408-464.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;5(1):44-56.
3. Riess O, Sturm M, Menden B, et al. Genomes in clinical care. *NPJ Genom Med*. 2024;9:20.
4. Struelens MJ, Ludden C, Werner G, Sintchenko V, Jokelainen P, Ip M. Real-time genomic surveillance for enhanced control of infectious diseases and antimicrobial resistance. *Front Sci*. 2024;2:1298248.
5. Gaye A, Marcon Y, Isaeva J, et al. DataSHIELD: taking the analysis to the data, not the data to the analysis. *Int J Epidemiol*. 2014;43(6):1929-1944.
6. Chen J, Qian F, Yan W, Shen B. Translational biomedical informatics in the cloud: present and future. *BioMed Res Int*. 2013;2013:658925.
7. Sunyaev A, Chorny D, Mauro C, Krcmar H. Evaluation framework for personal health records: Microsoft HealthVault vs. Google Health. In: *2010 43rd Hawaii International Conference on System Sciences*. IEEE; 2010: 1-10.
8. Banks D, Erickson JS, Rhodes M. Toward Cloud-based Collaboration Services. In: *USENIX Workshop on Hot Topics in Cloud Computing*. USENIX Association; 2009.
9. Rosenthal A, Mork P, Li MH, Stanford J, Koester D, Reynolds P. Cloud computing: a new business paradigm for biomedical information sharing. *J Biomed Inform*. 2010;43(2):342-353.
10. Liu F, Tong J, Mao J, et al. NIST cloud computing reference architecture. *NIST Spec Publ*. 2011;500(2011):1-28.
11. Chen YP, Paxson V, Katz R. What's New About Cloud Computing Security: Report No. UCBECS-2010-5. University of California, Berkeley. Accessed October 30, 2025. <https://core.ac.uk/outputs/24057828/>
12. Chokshi DA, Parker M, Kwiatkowski DP. Data sharing and intellectual property in a genomic epidemiology network: policies for large-scale research collaboration. *Bull World Health Organ*. 2006;84(5):382-387.
13. Baru C, Botts N, Horan T, Patrick K, Feldman SS. A seeded cloud approach to health cyberinfrastructure: preliminary architecture design and case applications. In: *2012 45th Hawaii International Conference on System Sciences*. IEEE; 2012:2727-2734.
14. Reid JG, Carroll A, Veeraraghavan N, et al. Launching genomics into the cloud: deployment of Mercury, a next generation sequence analysis pipeline. *BMC Bioinform*. 2014;15(1):30.
15. Aarestrup FM, Albeyatti A, Armitage WJ, et al. Towards a European health research and innovation cloud (HRIC). *Genome Med*. 2020;12(1):18.
16. Chentharas S, Ahmed K, Wang H, Whittaker F. Security and privacy-preserving challenges of e-health solutions in cloud computing. *IEEE Access*. 2019;7:74361-74382.
17. AbuKhousa E, Mohamed N, Al-Jaroodi J. E-health cloud: opportunities and challenges. *Future Internet*. 2012;4(3):621-645.
18. Mugonza R, Basaza-Ejiri AH. Issues affecting health research collaborations based on cloud computing. *Int J New Technol Res*. 2018;4(9):30-38.
19. Dong N, Jonker H, Pang J. Challenges in eHealth: from enabling to enforcing privacy. In: *Foundations of Health Informatics Engineering and Systems*. Springer; 2012: 195-206.
20. Popović K, Hocenski Ž. Cloud computing security issues and challenges. In: *Proceedings of the 33rd International Convention Mipro*. IEEE; 2010: 344-349.
21. Ranjan R. The cloud interoperability challenge. *IEEE Cloud Comput*. 2014;1(2):20-24.
22. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:n160.
23. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Med Res Methodol*. 2008;8(1):45.
24. Afgan E, Baker D, Batut B, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W537-W544.
25. Heath AP, Greenway M, Powell R, et al. Bionimbus: a cloud for managing, analyzing and sharing large genomics datasets. *J Am Med Inform Assoc*. 2014;21(6):969-975.
26. Dehnavieh R, Haghdoost A, Khosravi A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. *Health Inf Manag J*. 2019;48(2):62-75.
27. Ananthakrishnan R, Chard K, Foster I, Tuecke S. Globus platform-as-a-service for collaborative science applications. *Concurr Comput Pract Exp*. 2015;27(2):290-305.
28. Dahlquist JM, Nelson SC, Fullerton SM. Cloud-based biomedical data storage and analysis for genomic research: Landscape analysis of data governance in emerging NIH-supported platforms. *Hum Genet Genom Adv*. 2023;4(3):100196.
29. Libuit KG, Doughty EL, Otieno JR, et al. Accelerating bioinformatics implementation in public health. *Microb Genom*. 2023;9(7):001051.
30. Li R, Scott J, Rockhold F, et al. Moving data sharing forward: the launch of the vivli platform. *NAM Perspect*. 2018;8(11):1-4.
31. Mohammed-Rajput NA, Smith DC, Mamlin B, Biondich P, Doebbeling BN. OpenMRS, a global medical records system collaborative: factors influencing successful implementation. In: *Proceedings of the AMIA Annual Symposium*. American Medical Informatics Association; 2011;2011:960-968.
32. Fazen LE, Chemwolo BT, Songok JJ, et al. AccessMRS: integrating OpenMRS with smart forms on android. In: *Proceedings of MEDINFO 2013*. IOS Press; 2013;192:866-870.
33. Xia H, Asif I, Zhao X. Cloud-ECG for real time ECG monitoring and analysis. *Comput Meth Programs Biomed*. 2013;110(3):253-259.
34. Skolariki K, Avramouli A. The use of translational research platforms in clinical and biomedical data exploration. In: *GeNeDis 2016*. Springer International Publishing; 2017;938:301-311.
35. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401-404.
36. Ohno-Machado L, Bafna V, Boxwala AA, et al. iDASH: integrating data for analysis, anonymization, and sharing. *J Am Med Inform Assoc*. 2012;19(2):196-201.
37. Toga AW, Bhatt P, Ashish N. Global data sharing in Alzheimer's disease research. *Alzheimer Dis Assoc Disord*. 2016;30(2):160-168.
38. Atthey BD, Braxenthaler M, Haas M, Guo YK. tranSMART: an open source and community-driven informatics and data sharing platform for clinical and translational research. *AMIA Summits Transl Sci Proc*. 2013:6-8.
39. Afgan E, Chapman B, Taylor J. CloudMan as a platform for tool, data, and analysis distribution. *BMC Bioinform*. 2012;13(1):315.
40. Eric SD, Nicholas TKDD, Adiku Theophilus K. Bioinformatics with basic local alignment search tool (BLAST) and fast alignment (FASTA). *J Bioinform Seq Anal*. 2014;6(1):1-6.
41. Sahni S, Khanna A, Rodrigues JJPC. Analysis of biological information using statistical techniques in cloud computing. In: *Applications of Cloud Computing*. 1st ed. CRC Press; 2020:1-24.
42. Nguyen T, Shi WS, Ruden D. CloudAligner: a fast and full-featured MapReduce based tool for sequence mapping. *BMC Res Notes*. 2011;4:171.
43. Feng X, Grossman R, Stein L. PeakRanger: a cloud-enabled peak caller for ChIP-seq data. *BMC Bioinform*. 2011;12(1):139.
44. Jourden L, Bernard M, Dillies M-A, Le Crom S. Eoulsan: a cloud computing-based framework facilitating high throughput sequencing

- analyses. *Bioinformatics*. 2012;28(11):1542-1543.
45. Zhao S, Prenger K, Smith L, et al. Rainbow: a tool for large-scale whole-genome sequencing data analysis using cloud computing. *BMC Genom*. 2013;14(1):425.
 46. KG Thiruvilwamala. CareCloud: An Application for smart Personal Health Care Devices. Accessed May 18, 2026. <https://repositories.lib.utexas.edu/items/cb963338-b4c3-4eec-a370-c5e1924bddba>
 47. Doel T, Shakir DI, Pratt R, et al. GIFT-Cloud: a data sharing and collaboration platform for medical imaging research. *Comput Meth Programs Biomed*. 2017;139:181-190.
 48. Hoyt RE, Snider D, Thompson C, Mantravadi S. IBM Watson analytics: automating visualization, descriptive, and predictive statistics. *JMIR Public Health Surveill*. 2016;2(2):e157.
 49. Aggarwal M, Madhukar M. IBM's Watson analytics for health care: a miracle made true. In: *Cloud Computing Systems and Applications in Healthcare*. IGI Global; 2016: 117-134.
 50. Ko G, Kim P-G, Yoon J, et al. Closha: bioinformatics workflow system for the analysis of massive sequencing data. *BMC Bioinform*. 2018;19(1):43.
 51. Agapito G, Cannataro M, Guzzi PH, Marozzo F, Talia D, Trunfio P. Cloud4SNP: distributed analysis of SNP microarray data on the cloud. In: *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*. ACM; 2013: 468-475.
 52. Yung CK, Mihaiescu GL, Tiernay B, et al. Abstract 378: the cancer genome collaboratory. *Cancer Res*. 2017;77(13_Supplement):378.
 53. Prior FW, Clark K, Commey P, et al. TCIA: an information resource to enable open science. In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE; 2013: 1282-1285.
 54. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*. 2020;578(7793):82-93.
 55. Raney BJ, Cline MS, Rosenbloom KR, et al. ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res*. 2011;39(suppl_1):D871-D875.
 56. Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol (Pozn)*. 2015;1A:68-77.
 57. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O'Sullivan C. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res*. 2022;50(D1):D387-D390.
 58. Lee SM, Majumder MA. National institutes of mental health data archive: privacy, consent, and diversity considerations and options for improvement. *AJOB Neurosci*. 2022;13(1):3-9.
 59. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.
 60. Burgess DJ. The TOPMed genomic resource for human health. *Nat Rev Genet*. 2021;22(4):200.
 61. Sathe S, Ware J, Levey J, et al. Enroll-HD: an integrated clinical research platform and worldwide observational study for Huntington's disease. *Front Neurol*. 2021;12:667420.
 62. Brown E, Chahine L, Alonso D, et al. Parkinson's Progression Marker Initiative (PPMI) Online identifies people with features of prodromal Parkinson's disease through online questionnaires. Accessed May 18, 2026. <https://www.mdabstracts.org/abstract/parkinsons-progression-marker-initiative-ppmi-online-identifies-people-with-features-of-prodromal-parkinsons-disease-through-online-questionnaires/>
 63. Tryka KA, Hao L, Sturcke A, et al. NCBI's database of genotypes and phenotypes: dbGaP. *Nucl Acids Res*. 2014;42(D1):D975-D979.
 64. Toga AW, Neu SC, Bhatt P, Crawford KL, Ashish N. The global Alzheimer's association interactive network. *Alzheimers Dement*. 2016;12(1):49-54.
 65. McHugh CP, Clement MHS, Phatak M. AD Workbench: Transforming Alzheimer's research with secure, global, and collaborative data sharing and analysis. *Alzheimers Dement*. 2025;21(5):e70278.
 66. Barrett T. Gene expression omnibus (GEO). In: *NCBI Handb*. National Center for Biotechnology Information; 2013: 349.
 67. Conroy MC, Lacey B, Bešević J, et al. UK Biobank: a globally important resource for cancer research. *Br J Cancer*. 2023;128(4):519-527.
 68. Lappalainen I, Almeida-King J, Kumanduri V, et al. The European Genome-phenome Archive of human data consented for biomedical research. *Nat Genet*. 2015;47(7):692-695.
 69. Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. *Blood*. 2017;130(4):453-459.
 70. Zhang J, Baran J, Cros A, et al. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database*. 2011;2011:bar026.
 71. Burgin J, Ahamed A, Cummins C, et al. The European nucleotide archive in 2022. *Nucleic Acids Res*. 2023;51(D1):D121-D125.
 72. Kim S, Kim T, Lee S, et al. Cloud computing in population-scale genomics: Strategies, challenges, and future directions. *J Genet Med*. 2025;22(2):37-46.
 73. Franzoso E, Santorsola M, Lescai F. Rapid NGS analysis on google cloud platform: performance benchmark and user tutorial. *Clin Transl Sci*. 2025;18(11):e70416.
 74. Harnal S, Sharma G, Malik S, et al. Current and future trends of cloud-based solutions for healthcare. In: Tiwari, R., Koundal, D., Upadhyay, S. (eds) *Image Based Computing for Food and Health Analytics: Requirements, Challenges, Solutions and Practices*. Springer International Publishing; 2023: 115-136.
 75. EH Cloud. Health Data Hosting - Global Compliance. Accessed May 18, 2026. <https://www.euris.com/health-data-hosting/health-cloud-data-hosting/>
 76. Oshri I, Kotlarsky J, Willcocks L. Internet-based sourcing: cloud and crowdsourcing as delivery models. In: Oshri I, Kotlarsky J, Willcocks LP, eds. *Transformation in Global Outsourcing*. Springer International Publishing; 2024: 393-423.
 77. Giroud A, Puck J, Pühr H. World Investment Report 2025: international investment in the digital economy. *J Int Bus Policy*. 2026;9(1):146-152.
 78. Schatz MC. CloudBurst: highly sensitive read mapping with MapReduce. *Bioinformatics*. 2009;25(11):1363-1369.
 79. Jalili V, Afgan E, Gu Q, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res*. 2020;48(W1):W395-W402.
 80. Lau JW, Lehnert E, Sethi A, et al. The cancer genomics cloud: collaborative, reproducible, and democratized—a new paradigm in large-scale computational research. *Cancer Res*. 2017;77(21):e3-e6.
 81. Afgan E, Baker D, Coraor N, Chapman B, Nekrutenko A, Taylor J. Galaxy CloudMan: delivering cloud compute clusters. *BMC Bioinform*. 2010;11(Suppl 12):S4.
 82. Neu SC, Crawford KL, Toga AW. Sharing data in the global Alzheimer's association interactive network. *NeuroImage*. 2016;124:1168-1174.
 83. Charlebois K, Palmour N, Knoppers BM. The adoption of cloud computing in the field of genomics research: the influence of ethical and legal issues. *PLoS One*. 2016;11(10):e0164347.
- How to cite this article: Mugonza R, Ssembatya R, Basaza-Ejiri A. Cloud computing-based collaborations on health research: A systematic review of applications, barriers to adoption, and migration strategies. *Digit Med*. 2026;12:e26-00006. doi: 10.1097/DM-2026-00006