basepair

The Basepair Approach to Data Governance and Security

Bioinformatics Reinvented

We are reinventing bioinformatics from the ground up to address the ever-changing needs of today's genomics industry. Our domain knowledge of next-generation sequencing (NGS) and state-of-the-art Software-as-as-Service (SaaS) architectures is not only accelerating time to scientific insight, but it's also transforming the way genomics is supporting life-saving research.

- **Our Vision:** To unlock the potential of genomic data by helping every life sciences professional become a scientist and every scientist become a specialist. By empowering more people to understand the meaning of genomic data we can help accelerate more scientific breakthroughs as a community.
- **Our Platform:** Powers some of the world's leading biopharma and genomic technology startups, including Nkarta Therapeutics and Arima Genomics, as well as larger life sciences organizations like Vitrolife and Takara Bio. Similarly, we also support world-class research in academic medical centers, such as Harvard, MD Anderson, and Yale.

At its core, the Basepair platform offers researchers and clinicians a powerful way of understanding the underlying mechanisms of health and disease. Yet by its very personal nature, the value and sensitivity of genomic data means it must be protected with advanced security procedures. We ensure high levels of security & data privacy for our customers, having developed an advanced data governance and security strategy that sets us apart from most in the industry. This whitepaper focuses on how Basepair secures genomic data with its multi-layered, security-by-design approach to software development.

Data Flow within the Platform

To understand our approach to data governance, first it's important to understand how the platform accesses and analyzes genomic data.

Traditional commercial data sharing and analysis approaches involve the movement of sensitive data into centralized environments of hosted bioinformatics products. However, these approaches often provide overly functional user interfaces, don't scale cost effectively, and, most importantly, risk data security each time data is moved. To achieve ultimate security, data must remain within a customer's direct control. With Basepair, we provide that freedom.

We can configure a customer's settings to leverage the computing and storage resources they need within their own cloud account. With this architecture, both the data and analysis algorithms remain securely within the bounds and security firewalls of their secure cloud environment, The Basepair Platform assumes an Identity Access Management (IAM) role and interfaces with the customer's existing environment via a series of application programming interface (API) calls to spin up compute instances and execute read/write operations to the customer's cloud storage buckets.

We call this our Connected Cloud architecture (see *Figure 1*). This unique implementation in the bioinformatics industry not only removes 90% of compliance, security, and data privacy concerns, but it also enables our customers to control cloud costs while staying connected to other tools and resources in their cloud account. It's truly a win-win.

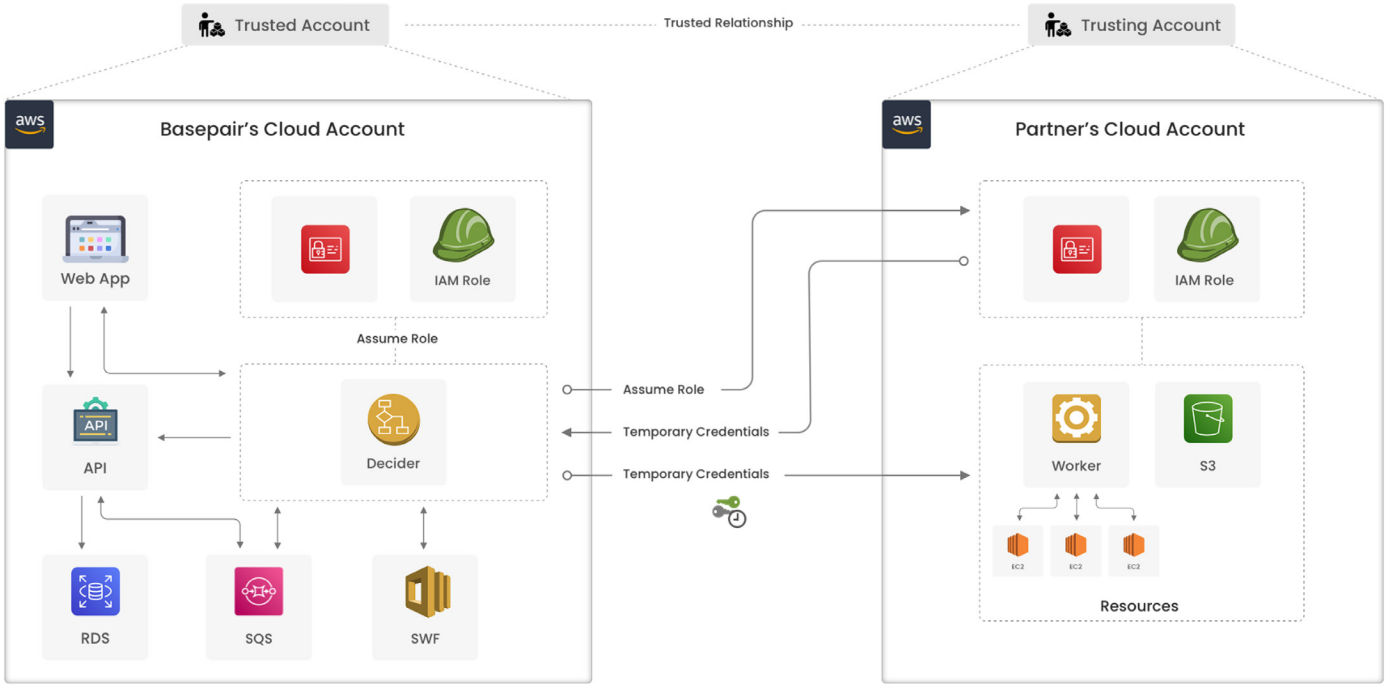

Figure 1. Connected Cloud architecture.

Accelerate Time to Insights without Compromising Security

Data management platforms must hold data securely and have industry-standard controls such as data encryption and the ability to track researcher/user activity. Data needs to be de-identified and encrypted both at rest and in transit. Basepair has designed an integrated system of internal controls, automated processes, monitoring, and risk assessment to establish a robust security plan based on these key areas: access, data, and privacy. Here's how it works.

Access

To restrict and monitor who accesses our platform and data, the platform includes the following safeguards.

User Authentication and Tiered Access

We ensure only authorized, verified users can access the platform through an email verification process. In addition, the platform enforces strict controls to ensure users have access only to data they're authorized to view. Authenticated users can specify permission by sharing specific projects, samples, or analyses with other users. Additionally, using identity management services like PingIdentity PingOne, the platform supports single sign-on functionality, enabling users within an organization to use their active directory accounts to log in. We also maintain compliance across multiple users in various locations. You can even roll back certain actions to avoid security and compliance issues that may have been caused by a user error.

Minimum Access Policy

We enforce a minimum access policy, meaning data access only to its intended owner or by explicit sharing consent. Authenticated users can view only the specific subset of data they own or data from projects, analyses, or samples explicitly shared with them. Basepair collaborates with clients to design solutions that segregate data, users, and projects across separate environments, so they can meet strict compliance requirements while retaining a highly performing, integrated system.

User Monitored Logging and Auditing

The platform has multi-layered monitoring capabilities to track and audit analyses and datasets within the research environment. These services monitor and log all user and cloud activity, including platform and data-specific information. Audit data can be provided to the client to enable them to proactively monitor data security in real time, to identify suspected unauthorized data access, data leaks, or anomalous activity. Any suspicious activity detected in the logs is reported directly to clients.

Data

Our Connected Cloud capability acts as a form of Federated Architecture, meaning customers retain full security over their data at all times. All data and associated analyses remain within the bounds and security firewalls of the customer's environment. This is an increasingly important approach to minimize data egress and movement, enabling access to and analysis of genomic data when it can't be exported for permissions, data residency, or other reasons.

Encryption

The platform also has strong data encryption standards, using SSL/TLS as a baseline. Data is encrypted at rest (eg, when data is in storage), in transfer (eg, when data is moving between storage buckets and compute machines), and during analysis. Data can only be de-encrypted by authenticated staff, and the security network imposes additional constraints regarding which users can access, view, or edit encrypted files.

Data Integrity

All genomic data is encrypted to comply with HIPAA, the GDPR Security Rule, and Secretary of Health and Human Services guidelines. We even encrypt all of the user's data, including usage and other metadata our architecture splits from genomic data for multiple layers of protection against breaches and other threats.

Virtual Private Cloud (VPC)

Customers can ensure their data source integrity and security is protected, as the platform has a one-way ingestion to create a separate pool of analysis-ready data, while the original data source is not changed. Further, data doesn't persist beyond purposes directly relevant to the research, as the downstream compute environment is spun up only for computing and is destroyed immediately after. In addition, IDs are used to de-identify personal health information from any biomaterial data for an individual sample.

Secure API

If a customer wants to upload genomic data into the Basepair hosted service, they can rest assured knowing it's stored in a highly secure Virtual Private Cloud (VPC) powered by Amazon Web Services (AWS) with built-in HIPAA and GDPR compliance.

Our end-to-end data encryption allows developers to upgrade or build new enterprise apps in simple environments using tools, such as Docker, with our secure API. The platform also maintains and monitors highly performance operations at a system-wide level, including the following:

- Business Continuity and Disaster Recovery Plan for security incident response. As part of our continuous improvement of security and incident response planning, we complete an annual review of cybersecurity policies to stay current on changing best practices and legislation.
- Vulnerability and Penetration Testing - Basepair operates continuous vulnerability scanning and has a stringent testing regime, including annual penetration tests performed by accredited, independent providers. With continuous 99.98% or better average uptime, we are always on the lookout for the latest cyber threats. In addition, we regularly test and patch vulnerabilities as we identify them.

Privacy Measures

Compliance

At Basepair, we have meticulously architected our platform to meet or exceed HIPAA and GDPR specifications. Although the platform itself is not CAP/CLIA or 21 CFR Part 11 certified, Basepair complies with GxP regulations enforced by the US Food and Drug Administration (FDA), which makes the compliance process easier for our customers and helps them create a certified solution. Additionally, we follow dbGaP Security Best Practices for handling genomic data, working proactively with clients to comply with sensitive data requirements and ensure organizations meet and exceed industry standards.

Staff & Vendor Requirements

You can trust us to keep your data secure. All Basepair staff undergo background checks, and we require all staff to complete security training when they're onboarded, and on an annual basis. Vendors processing confidential data are subject to supplier risk assessments and due diligence processes.

Available in AWS Marketplace

aws PARTNER Qualified Software

AWS Partner Network and Foundational Technical Review (FTR)

We have passed the AWS Foundational Technical Review (FTR) to achieve Qualified Software status as a member of the AWS Partner Network. The FTR provides crucial support for data governance and security, offering a comprehensive assessment of an organization's AWS environment. This ensures the environment aligns with AWS best practices and industry standards, so customers can trust their data is safe.

The FTR evaluates various aspects of the AWS infrastructure, including network architecture, identity, access management, data protection, monitoring, logging, and compliance. By conducting this review, AWS helps customers identify potential security vulnerabilities, ensure proper access controls, and establish strong data protection measures.

Through the FTR, AWS assists in establishing a secure foundation for data governance by implementing encryption, data classification, and access controls. It helps define data governance policies, establish data ownership, and implement privacy and compliance controls. By adhering to industry-leading security standards and best practices, the FTR strengthens data governance and ensures sensitive information stays protected.

The FTR also enables organizations to monitor their AWS environment. It helps establish logging and monitoring capabilities to detect and respond to security incidents promptly. The review assesses the organization's incident response processes, ensuring they're well-defined and tested. Overall, the AWS Foundational Technical Review plays a vital role in supporting data governance and security.

Summary

At Basepair, we understand genomic data is invaluable, yet its scale and sensitivity can pose unique challenges for data sharing.

That's why we've created a sustainable and resilient bioinformatics platform for companies of all sizes to safeguard their scientific data. We continually review our security policies to ensure customers can meet changing regulatory and legal requirements.

The steps we've taken to go above and beyond to ensure scientific data stays secure is a testament to our commitment to making genomic data analyzable for life-saving research.

Contact us!

sales@basepairtech.com

www.basepairtech.com

+1 (347) 428-9399