Element Biosciences

ElemBio™ Cloud Data Protection

Secure and transparent computing in a flexible data ecosystem fit for the Element AVITI™ System

Introduction

Genomics research requires safe management of sequencing data to reduce risk to privacy and the misuse of genetic information. ElemBio Cloud is an online platform that functions as a direct extension of Element AVITI Systems. ElemBio Cloud allows remote management and monitoring of runs and instrument fleets in real-time, and includes a rich suite of data analysis tools. With ElemBio Cloud, you can orchestrate how you send sequencing data to your preferred environment, and truly own and manage your data with full control (Figure 1). If you prefer to not allow Element Biosciences to access any sequencing data, then do not subscribe to ElemBio Catalyst™.

This document discloses how ElemBio Cloud helps operate online systems, including interactions with telemetry and cloud services. From product conception through commercialization, Element Biosciences implements data security and privacy protections, and is committed to complying with applicable global data privacy regulations. For more information, see our privacy policy at elementbiosciences.com/legal/privacy.

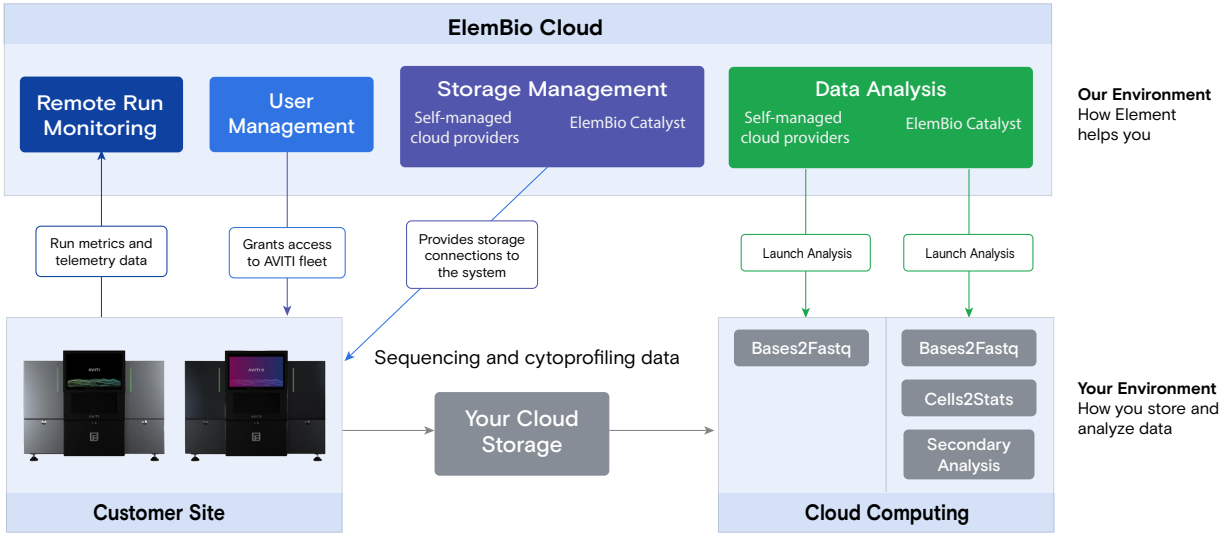

Figure 1. ElemBio Cloud provides a framework for customized data analysis and allows you to work in a preferred environment with as-needed assistance from Element.

Element Biosciences

Data control and ownership

Despite many advancements, sequencing data analysis remains complex, opaque, and an impediment to results. To quickly facilitate analysis and get results, ElemBio Cloud offers scalable cloud-based solutions that help you analyze the AVITI System data. The flexible ecosystem of Element solutions allows you to control your data exactly how you want to. Your data, your way!

If you would like help with an analysis solution, Element provides **ElemBio Catalyst** as a subscription service. With ElemBio Catalyst, Element builds, manages, and hosts a cloud analysis solution on your behalf. Your data is stored at Element in storage buckets that are completely dedicated to you. Data is stored in your required region. If you don't want to use ElemBio Catalyst, you can manage your data with costs operated by your preferred cloud provider. Element supports AWS, Google Cloud, and DNAnexus.

Online benefits versus offline

AVITI Operating Software (AVITI OS) controls system operations and includes telemetry for proactive support and maintenance, along with online and offline modes. An AVITI System in online mode automatically connects to telemetry and ElemBio Cloud and accesses cloud storage, self-managed software updates, informatics support, and other conveniences (Table 1).

For labs that prefer to maintain all data locally, offline mode supports a USB drive or local network server for data storage. Operating an AVITI System in offline mode turns off telemetry, which impacts the Element response time to provide support. Offline instruments benefit from some proactive maintenance and support, but with limitations on capabilities and other services.

Service or Feature	Online Mode	Offline Mode
Telemetry	Yes	No
Cloud storage connection	Yes	No
ElemBio Catalyst	Yes	No
Remote run monitoring	Yes	No
Account management[a]	Yes	Yes

[a] For example, download a product key or submit a support request.

Table 1. Comparison of online and offline modes in AVITI systems

Secure framework for sequencing data analysis

Security starts with the AVITI System, which generates and processes data for primary analysis. AVITI OS combines several microservices on Ubuntu Core 20.04 LTS to deliver a highly controlled and secure operating system (OS) for embedded devices.[1] These microservices run in snaps, which are self-contained applications that run in fully isolated (sandboxed) environments, and prohibit unauthorized scripts and executables from being installed or ran on the the AVITI System.[2]

An onboard field-programmable gate array (FPGA) processes images for primary analysis and enables compact, low-power computation that sustains tremendous throughput. Data is processed throughout a run and allows a host of information to be available in real-time through AVITI OS or ElemBio Cloud. For more information on network architecture and security, see the ***AVITI System Site Prep Guide (MA-00007)*** and ***AVITI24 System Site Prep Guide (MA-00052)*** at elementbiosciences.com/resources.

Element Biosciences

Controlled access via permissions

The AVITI System is a secure, protected resource. ElemBio Cloud includes the ability to add and remove system access on a user-by-user basis. To further manage who controls each system, labs can connect ElemBio Cloud to an authentication provider through single sign-on (SSO).

Users can use their browser to access ElemBio Cloud, where they can plan and monitor runs and view integrations and flows. Role-based access separates administrators from team members and allows only administrators to manage integrations, flows, and team members. To ensure quick addition and removal of access, user control is instant.

Telemetry for proactive maintenance

Reliable system operation optimizes usage and maximizes uptime. To promote reliability, the AVITI System employs telemetry. Telemetry is the automatic measurement and over-the-air (OTA) transmission of performance data from remote sources. Sensors that are embedded throughout the AVITI System measure performance and operation to generate system health data that is sent to Element. Telemetry also allows Element Support to review sensor data, notifications, and AVITI OS metrics to identify potential problems and provide proactive service and support. Telemetry data is protected with permissions that allow only Element teams with approved clearance.

Telemetry data is carefully curated to protect sensitive information and **does not include any sequencing or genomic data** output from an AVITI run. Thus, telemetry **does not** collect the following data:

- Run manifests
- Sequencing run data (base calls) or cytoprofiling run data
- Sample names, well names, and run descriptions
- Personal data and personally identifiable information (PII)
- Protected health information (PHI)

Telemetry safely accesses only the information necessary to identify potential problems (Table 2).

Telemetry Data	Collected Information	Benefit
Instrument information	Serial number, instrument ID, and instrument name	Identifying the instrument generating telemetry data
Run information	Number of cycles per read, consumable information, and the following run information: • Name • ID • Side (A or B) • Type (sequencing, cytoprofiling, or wash) • Start and end dates and times • Progress updates and status (e.g., completed)	Identifying the run generating the telemetry data and experimental variables
Hardware metrics	Motor, fan, and laser performance	Assessing the condition of hardware components
Software metrics	Memory, software and firmware versions, and CPU performance	Assessing the performance of software components
Primary analysis metrics	Q30, error rates, index assignment metrics without sample names, assigned rates, mismatch rates, and cell assigned counts	Assessing overall run health for quality control (QC) purposes
System logs	Power cycle times, errors, internal communications, internal services statuses	Analyzing factors that might contribute to system performance
Run logs	Recipe execution, timing of run steps, and communications between software, firmware, and hardware	Analyzing factors that might contribute to runs performance

Table 2. Data that telemetry collects

Element Biosciences

Protecting run data during transfer and storage

Run data generated by an AVITI System in online mode is stored off-instrument in local storage or your cloud account. A storage connection links AVITI OS to a bucket in the cloud account, allowing the transfer of data to cloud storage. Through ElemBio Cloud, labs can set up and manage storage connections. During each run setup, the system operator selects a storage connection or accepts the default storage connection for the organization.

Run data and run metrics become visible as they are generated and transfer directly from the instrument to cloud storage throughout the run. When the run is complete, all data is available in a bucket. The authentication method that secures the transfer of data is based on on the storage connection type (Table 3). Labs that use an Amazon Web Services (AWS), DNAnexus, or Google Cloud Storage (GCS) connection retain full ownership over run data, which never passes through an Element account.

ElemBio Catalyst uses Amazon Simple Storage Service (S3) to maintain buckets on your behalf. The bucket remains separate from other customers and access is restricted to a limited number of Element employees for maintenance purposes only. Otherwise, Element staff cannot access the bucket or the data it contains and the ElemBio Cloud infrastructure keeps data streams separate. ElemBio Catalyst incorporates data security and privacy protections, including those from our provider, AWS, in accordance with applicable data privacy regulations.[3–5]

Storage Connection	Bucket	Authentication
AWS	Amazon S3	Role-based or secret key authentication through Identity and Access Management (IAM)
DNAnexus	Varies	API key authentication
ElemBio Catalyst	Amazon S3	Role-based or secret key authentication through IAM
GCS	Cloud Storage	Secret key authentication through a keyed-hash message authentication code (HMAC)

Table 3. Authentication methods for each type of cloud storage connection

Unrestricted secondary analysis via integrations

ElemBio Cloud streamlines your secondary analysis with premium workflow add-ons in ElemBio Catalyst or connections to third-party analysis platforms. After a sequencing run, Bases2Fastq software converts the bases files into FASTQ files that are ready for secondary analysis. Then, subscription-based premium workflows use tools, such as Sentieon and NVIDA Parabricks, to enable whole genome or enrichment secondary analysis. Analysis tools that are used in ElemBio Catalyst are secure and sequencing data is not visible or recorded in any manner outside of ElemBio Cloud. For more information on ElemBio Catalyst premium workflows, see *Secondary Analysis for Sequencing Workflows*.

If you do not subscribe to ElemBio Catalyst, then secondary analysis can be performed in a third-party compute platform that you choose and configure with your IT department (Figure 2). Labs can manually configure the compute platform or use a preconfigured and verified analysis flow from Element. Partnerships allow labs to establish the ideal combination of cloud computing components. For compatible partners, visit elementbiosciences.com/products/elembio-cloud.

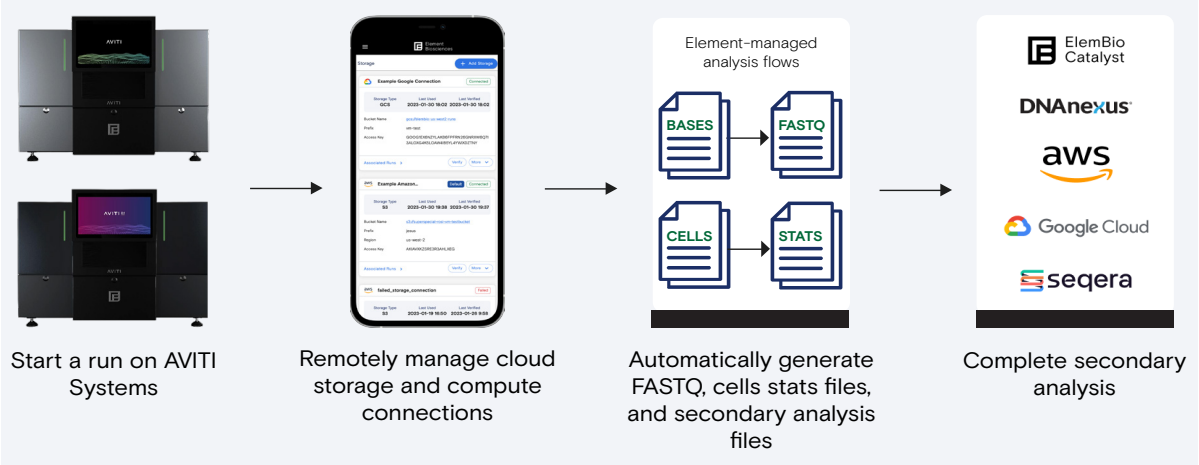

Figure 2. AVITI integrates into an end-to-end solution for processing data in a contained environment. Whether you use ElemBio Catalyst or a partner service, ElemBio Cloud facilitates FASTQ file generation and secondary analysis.

Orchestration that keeps data in your environment

Unless you subscribe to ElemBio Catalyst, Element doesn't receive or view sequencing data. This effectively restricts the role of Element to data orchestration, which is an automated process that makes cloud data available for analysis. Labs grant Element permission to operate in their environment and specify which actions Element can perform. For example, you can set up the ability to launch Bases2Fastq on Amazon HealthOmics from your storage location to demultiplex and convert bases files into FASTQ files.

Federated analysis allows data analysis to occur in your account. A set of local computations share only the information that is needed for processing, and leaves data secure. This method allows Element to bring operations to your account and perform actions on your behalf without Element accessing the data.

Conclusion

ElemBio Cloud provides an increasingly powerful platform that is founded on the principle that your sequencing data is your own. If you would like Element to help you with an analysis solution, ElemBio Catalyst is available as a subscription service. Element is continuing to expand developer tools, fleet management capabilities, and integrations in a safe, secure environment that supports where and how you want to work.

By leveraging partnerships and technological advancements such as federated analysis, ElemBio Cloud enables simpler, more accurate analysis with greater privacy. From a system fortified with an Ubuntu Core OS and permissions-based storage connections through orchestrated secondary analysis with ElemBio Catalyst, ElemBio Cloud promotes secure data management that gets any lab to an ideal solution with full transparency.

References

1. Rule4, Ubuntu Core Cybersecurity Analysis Technical White Paper, (2020): https://canonical.com/blog/ubuntu-core-a-cybersecurity-analysis.
2. "Introduction to snaps," Canonical, accessed June 19, 2023, https://ubuntu.com/core/services/guide/snaps-intro.
3. "AWS Cloud Security," AWS, accessed December 20, 2023, https://aws.amazon.com/security/.
4. "Privacy Policy," Element Biosciences, accessed January 11, 2024, https://www.elementbiosciences.com/legal/privacy.
5. "Terms of Use," Element Biosciences, accessed January 29, 2024, https://www.elementbiosciences.com/legal.

Element Biosciences
Telephone: 619.353.0300
Email: info@elembio.com