Cennelly Consulting
Information Security Management

GenRocket Test Data Automation

Security Controls Overview

Published May 2020

Table of Contents

Executive Summary

Thorough Quality Assurance testing is critical to ensure applications function as intended. This is particularly important for applications that process critical, confidential data, such as credit cards and personal information. Developers cannot use real data for testing and need the ability to generate test data that mimics the features of the data it will eventually process.

GenRocket's service offering enables its clients to create custom templates, called **Scenarios**, that are uploaded to the GenRocket cloud platform to generate synthetic test data for GenRocket clients that is downloaded into the clients' environment. This data may use industry standards (such as credit card number formats, or health care codes) or other models designed by the customer.

It is important to understand that the type of data that is hosted in the GenRocket cloud is **not** subject to regulations regarding Personally Identifiable Information (PII), Personal Health Information (PHI), credit card data, or other confidential information. Compliance requirements such as PCI, DSS and HIPAA require that these data types be de-identified before they can be used in testing, which is the purpose of generating synthetic test data.

At most, an organization may view their custom scenarios as Intellectual Property, but at no time is any data, subject to regulation, hosted in or transmitted through the GenRocket Cloud or infrastructure. This overview document describes controls that are in place to protect the confidentiality, availability and integrity of data in the GenRocket Cloud.

GenRocket business practices and technical controls are in line with industry best practices for the service they are offering. GenRocket is partnered with a third-party hosting provider which has achieved and maintained ISO 27001 certification to deliver back-end hosting services for the GenRocket platform.

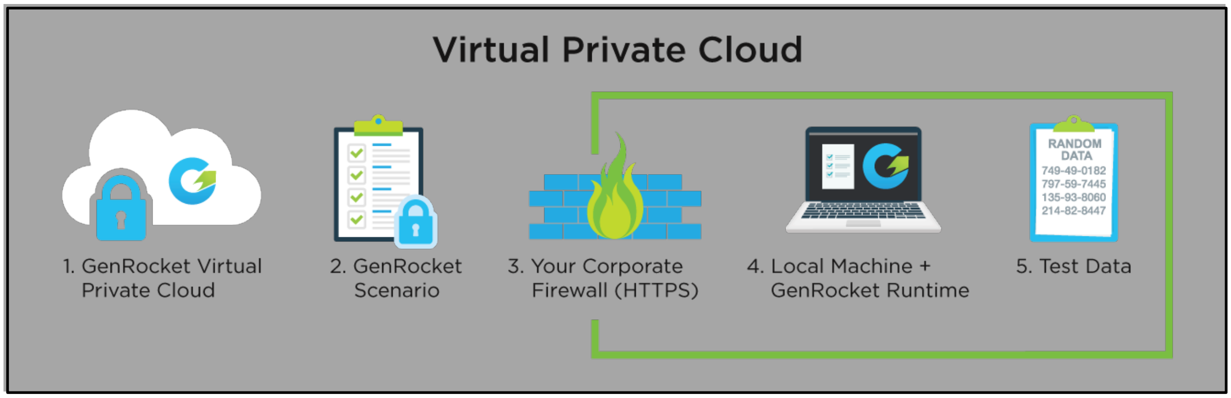
The GenRocket platform has excellent controls to preserve the confidentiality, integrity, and availability of the type of data contained in its environment. GenRocket does not store sensitive customer information in the system and all test data is generated within the customer's secure environment. Most importantly, synthetic test data by design is not subject to compliance regulations defining confidential information.

The GenRocket Ecosystem

Every aspect of GenRocket's platform was designed with security in mind to protect the confidentiality, integrity, and availability of clients' data.

GenRocket Hosting Models

GenRocket offers three hosting models: **Virtual Private Cloud**, **Dedicated Private Cloud**, or **On Premise**.



The **Virtual Private Cloud** hosts the GenRocket Runtime environment in GenRocket's secure Cloud environment. This is the solution preferred by GenRocket's smaller customers who do not have the infrastructure or staff to manage their own environment. Synthetic test data is generated in the GenRocket Secure Cloud environment where it can be retrieved by customers to be used in their private development environment. The Virtual Private Cloud solution provides multi-tenant hosting.

The **Dedicated Private Cloud** hosts the GenRocket Runtime environment in a dedicated GenRocket Secure Cloud environment. This is the solution preferred by larger customers who require their data to be segregated. Synthetic test data is generated in the Dedicated GenRocket Secure Cloud environment where it can be retrieved by customers to be used in their private development environment. The Dedicated Private Cloud solution provides single-tenant hosting.

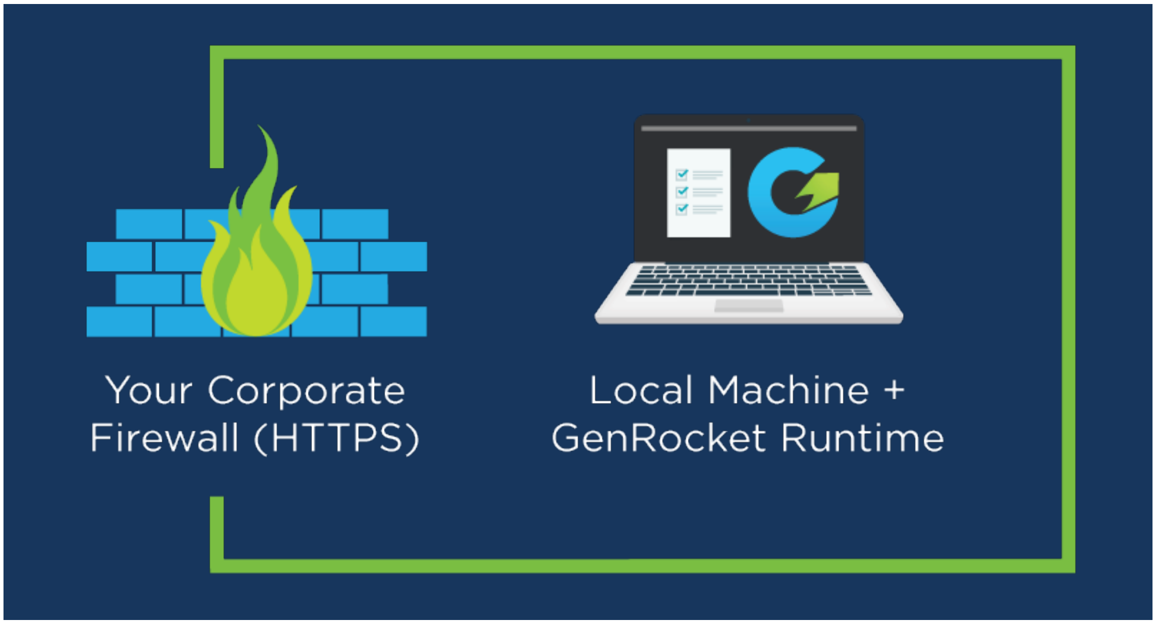
The **On-Premise solution** hosts the GenRocket Runtime in the customer's environment behind their firewall and managed by their staff. This solution requires that the client provides staff and IT resources to maintain the infrastructure. Test data is generated in the GenRocket Runtime environment and hosted on a server owned and managed by the customer.

It is important to note that at no time is any confidential data that is subject to regulatory compliance frameworks contained in the GenRocket Runtime environment or the GenRocket Virtual Cloud environment.

GenRocket Cloud and Server Infrastructure

The **GenRocket Cloud** and **Server Infrastructure** is hosted in an environment that is ISO 27001 Certified. All GenRocket communications are through secured socket layer using HTTPS in the client environment.

The GenRocket Runtime


Your Corporate Firewall (HTTPS)
Local Machine + GenRocket Runtime

The GenRocket Runtime is a secure Java program that executes the encrypted GenRocket Scenarios to generate test data in the customer's secured, firewalled environment. All GenRocket Runtime Java Archives (JARs) are validated with a checksum.

Data Types

Sensitivity of data is an important factor to consider when evaluating security controls. No organization wants to be in the headlines for a large data breach. For both the **Virtual Private Cloud** and the **Dedicated Private Cloud** solutions, it is important to understand that the type of data that is hosted is **not** subject to regulations regarding Personally Identifiable Information (PII), Personal Health Information (PHI), credit card data, or other confidential information. Compliance requirements such as PCI and HIPAA require that these data types be de-identified before they can be used in testing. Appendix 1 contains a summary of some relevant compliance requirements and Appendix 2 provides a description of how HIPAA data is de-identified. Appendix 3 provides an example of how a company may classify their data.

Data used in the GenRocket environment falls into the following categories: **Scenarios**, **Synthetic Test Data**, and **Published Data Standards**.

Scenarios

Scenarios are instruction sets that contain no data, modeled by customers in the GenRocket Cloud and downloaded to a computer in the customer's secured environment.

- Scenarios are encrypted
- Only authenticated users can run Scenarios
- Synthetic test data is generated on premise in the customer's secure environment

Data Classification: While Scenarios contain no data, they could be considered the Intellectual Property of the customer, since they are created by the customer. In a Data Classification scheme, this would be considered "Proprietary" (Internal Use Only). GenRocket provides security controls that are consistent with industry guidance on protection of Proprietary information.

Synthetic Test Data

Synthetic test data is mocked up data that aligns with a particular standardized format and is used to test algorithms for robustness during application development.

- Synthetic test data is generated in the customer's secured environment.

Data Classification: Synthetic test data itself could be treated as Public, since it is not associated with other information that would make it valid. GenRocket provides protections consistent with "Proprietary" data to ensure the integrity of the data used in application development.

Published Data Standards

Published Data Standards provide a common ground so that applications from different vendors can interact seamlessly. These standards are often used when generating synthetic test data.

Data Classification: By definition, Published Data Standards are Public. GenRocket provides protections of Scenarios and Synthetic Test Data consistent with "Proprietary" data to ensure the integrity of the data used in application development.

Summary

Every aspect of the GenRocket system has been designed from the start with security in mind to preserve the confidentiality, integrity, and availability of data. GenRocket does not store sensitive customer information in the system and all test data is generated within the customer's secure environment. Most importantly, synthetic test data by design is not subject to compliance regulations defining confidential information.

CFennelly Consulting
Information Security Management

CFennelly Consulting, LLC is a partnership co-managed by Carole Fennelly, an Information Security Management consultant in the Greater NYC area. Carole has over 35 years of hands-on experience in the Information Security and Technology fields and has authored several industry-standard security benchmarks based on her extensive experience in operating system platforms and security practices. As a consultant, Carole has defined security strategies and developed policies and procedures to implement strategies at numerous Fortune 500 clients in the NYC area. Carole maintains the following compliance certifications:

Certified HIPAA Professional (CHP)

License 201-002829

Certification Date Dec 2015 – Dec 2021

Certified Information Security Manager (CISM)

License 1220818

Certification Date Sep 2012 – Present

Certified Security Compliance Specialist (CSCS)

License 401-0000776

Certification Date Oct 2015 – Dec 2021

Appendix 1: Compliance Data Security Requirements

Compliance Requirement	Protected Data Type	Targeted Entities
Payment Card Industry Data Security Standard (PCI DSS)	• cardholder data • sensitive authentication data contained in a payment card's chip or magnetic stripe, including the 3-4 digit card verification code or value printed on the front or back of the payment card • personally identifiable payment card data	Merchants, Credit card processors, Credit card issuing banks, Credit card service providers
https://www.pcisecuritystandards.org/document_library		
Health Insurance Portability and Accountability Act (HIPAA) and Health Information Technology for Economic and Clinical Health (HITECH) Act	Patient Health Information (PHI) of U.S. Citizens. PHI is health information, such as illnesses and treatments, combined with identifying information about a specific individual, such as: name, address, social security number or other demographic data that would identify the individual associated with the illness or treatment. Health information that cannot be used to identify an individual is not PHI.	Organizations that handle patient health information of U.S. citizens, including healthcare providers, pharmacies, Pharmacy Benefit Managers, and Business Associates of any of the above.
https://www.hhs.gov/hipaa/index.html NIST Mapping to HIPAA Security Rule: https://www.hhs.gov/sites/default/files/nist-csf-to-hipaa-security-rule-crosswalk-02-22-2016-final.pdf HIPAA Security Rule Combined text (relevant section starts on page 62) https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf U.S. Department of Health & Human Services Breach Portal: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf		

Gramm-Leach-Bliley Act (GLBA)	Nonpublic Personal Information of U.S. Citizens	Financial institutions
http://www.ftc.gov/privacy/glbact/glbsub1.htm http://www.ftc.gov/privacy/privacyinitiatives/glbact.html		
Massachusetts Privacy Act **21 CMR 17.00**	Personally Identifiable Information (PII) of Massachusetts residents	Companies that handle personal information of Massachusetts residents
https://www.mass.gov/regulations/201-CMR-17-standards-for-the-protection-of-personal-information-of-residents-of-the		
General Data Protection Regulation (GDPR)	Personally Identifiable Information (PII) of European Union and the European Economic Area citizens	Companies that handle personal data of citizens in the European Union and the European Economic Area
https://gdpr-info.eu/		
California Consumer Privacy Act (CCPA)	Personally Identifiable Information (PII) of California residents	Companies that handle personal information of California Residents
https://oag.ca.gov/privacy/ccpa		
Fair Credit Reporting Act, **15 U.S.C. § 1681** **(FCRA)**	Consumer Credit Information of U.S. Citizens, including social security numbers, check writing history, medical records, and rental history records	Consumer Reporting Agencies, credit agencies
https://www.consumer.ftc.gov/articles/pdf-0096-fair-credit-reporting-act.pdf		

Appendix 2: De-Identifying HIPAA Data

List of 18 PHI Identifiers

The HIPAA Privacy Rule lists the following18 identifiers that characterize PHI:

1. Names
2. All geographical subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code, if according to the current publicly available data from the Bureau of the Census: (1) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
3. All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older
4. Phone numbers
5. Vehicle identifiers and serial numbers, including license plate numbers
6. Fax numbers
7. Device identifiers and serial numbers
8. Electronic mail addresses
9. Web Universal Resource Locators (URLs)
10. Social Security numbers
11. Internet Protocol (IP) address numbers
12. Medical record numbers
13. Biometric identifiers, including finger and voice prints
14. Health plan beneficiary numbers
15. Full face photographic images and any comparable images
16. Account numbers
17. Any other unique identifying number, characteristic, or code (note this does not mean the unique code assigned by the investigator to code the data)
18. Certificate/license numbers

De-identifying HIPAA Data

Once data has been de-identified, it is no longer subject to HIPAA controls. Data is considered to be “de-identified” if all of the 18 listed PHI identifiers have been removed and none of the remaining information could be used to identify an individual either alone or in combination with other information.

Appendix 3: Data Classification Example

Many organizations use a data classification method based on risks to the Confidentiality, Integrity, and Availability of the data. This is often referred to as a CIA rating. The method used for Data Classification varies based on the type of organizations, but usually fall into one of the following categories:

PUBLIC

This information has been specifically approved for public release by the Marketing/Business department. Unauthorized disclosure of this information is not materially harmful to the company its customers, or business partners.

Examples: Marketing brochures and materials, [COMPANY] web site content.

Authorization: Disclosure of the company information to the public requires the specific permission of [fill in], or standard precedent of publicly distributing this information.

Access Control: No authentication is required to view the data; but the integrity of the data must be protected from tampering. Access to the data may be logged, although this is not required.

PROPRIETARY (Internal Use Only)

This information is intended for use within the company only. Unauthorized disclosure of this information to external unauthorized recipients is not allowed and violation or risk to the company may result in legal action or damages.

Examples: Company organization charts, sensitive reports, and internal electronic mail messages not explicitly intended for external distribution.

Authorization: Disclosure of company information outside the company requires specific permission. Disclosure within the company is permitted without explicit permission from the information owner unless classified as CONFIDENTIAL.

Access Control: Authentication is required to access the internal network. All transmission of proprietary information to the external network must be traceable through logging mechanisms.

Examples: Employee personal information (such as social security numbers), customer transaction account information, and credit reports.

Authorization: Decisions about the provision of access to this information or dissemination of this information must be cleared through the [fill in], with consultation from Legal and Human Resources as appropriate.

Access Control: If electronically stored, a hardware or software mechanism must be in place to ensure access is granted to authorized users only. This mechanism must support the ability to trace access over a period of time.

References

PCI Data Do’s and Don’ts

https://www.pcisecuritystandards.org/pdfs/pci_fs_data_storage.pdf

Credit Card Formats

https://www.freeformatter.com/credit-card-number-generator-validator.html

NIST Special Publication 800-122

Confidentiality of Personally Identifiable Information

https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf

Methods for De-identification of Personal Health Information in Accordance with HIPAA

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

Code List for Certain Designated Health Services

https://www.cms.gov/Medicare/Fraud-and-Abuse/PhysicianSelfReferral/List_of_Codes