**Test Data Generation**

Liz Hinz

CSD370-A339: Secure Software Development

Professor Nathan Braun

May 4, 2025

The methods GeeksforGeeks provides for generating test data are manual test data generation, automated test data generation, back end data injection approach, and third-party tool. Generating manual test data means a developer or tester manually creates the test based on experience and resources (GeeksforGeeks, 2025). The best time to manually generate test data is when the tester is experienced, and other tools are unnecessary. Automated test data is generated automatically, making it faster to execute more data (GeeksforGeeks, 2025). The best time for an automated test is when you want fast and accurate results. Back end data injection involves SQL query injection based on test cases (GeeksforGeeks, 2025). Back end data injection is good when wanting to save time, and if testers have less experience. A third-party tool generates test data based on requirements and is highly customizable (GeeksforGeeks, 2025). A good time to use this method is when the budget allows for these types of expenses and when it is suitable for the environment. The article mentioned there are tools, but did not provide or recommend any tools available for generating the data using any of the four given methods.

https://www.geeksforgeeks.org/approaches-for-test-data-generation-in-software-testing/#

Keployio on the Medium website provides three ways to generate test data. Those are manual test data generation, automated test data generation, and production data cloning. Although time-consuming and potentially filled with human errors, manual test data can be simple and practical for more minor testing scales (Keployio, 2024). A good time for manual test data generation is when testing requirements are clearly defined. Automated test data generation is when test data is generated automatically while also being quick and accurate (Keployio, 2024). The best time for automated tools is when time needs to be saved and labor efforts need to be reduced. Production data cloning involves cloning a program or production environment for testing (Keployio, 2024). An excellent time for production data cloning is when a system is close to the deployment date. Some tools mentioned are Mockaroo, DataGenerator, Keploy, and SQL Data Generator.

https://medium.com/@keployio/test-data-generation-an-essential-guide-01832ca10c0f

Khullar on the Medium website provides different ways of generating test data through artificial intelligence. Data can be implemented through data synthesis, data masking and anonymization, and data augmentation (Khullar, 2024). Data synthesis creates real-life scenarios for testing (Khullar, 2024). A good time to use this is when testing user usage and input. Data masking and anonymization replace sensitive information (Khullar, 2024). A good time to use this is to ensure sensitive information is never exposed. Lastly, data augmentation goes off existing data for testing (Khullar, 2024). A good time to use this is to check how existing systems and forms work. There are some tools available like Faker, SynthDet, Mockaroo, Google DLP, Apache Nifi, and NLPAug (Khullar, 2024).

https://medium.com/@amitkhullaar/ai-driven-test-data-generation-52c3747dcb91

**References**

GeeksforGeeks. (2025, March 1). *Approaches for Test Data Generation in Software Testing*. GeeksforGeeks. https://www.geeksforgeeks.org/approaches-for-test-data-generation-in-software-testing/#

Keployio. (2024, October 11). *Test Data Generation: An Essential Guide - Keployio - Medium*. Medium. https://medium.com/@keployio/test-data-generation-an-essential-guide-01832ca10c0f

Khullar, A. (2024, February 27). *AI-Driven Test Data Generation*. Medium. https://medium.com/@amitkhullaar/ai-driven-test-data-generation-52c3747dcb91