



Service Definition - Synthetic Test Data Generation

Date : 30/01/2026

Version : 1.0



1 Overview

Build unshakeable confidence in testing critical citizen services with data that's realistic, compliant, and ready when you need it. 2i's award-winning Synthetic Test Data Generation service provides independent assurance that digital services are thoroughly tested without exposing production data in test environments. We enable public sector organisations to meet data protection commitments with confidence while protecting citizen privacy by generating test data that behaves like the real thing same patterns, same edge cases, same statistical properties without containing a single real citizen record, getting it right first time, every time.

Our value-for-money framework drives measurable improvements in testing effectiveness, delivery efficiency, and compliance confidence while demonstrating responsible use of public funds. By removing the dependency on production data, we help reduce operational risk to citizen-facing services while delivering auditable value for money within the first 3 months of engagements. Organisations gain the ability to test comprehensively without the constraints that typically delay delivery or compromise data protection.

This approach enables delivery teams to move faster with reduced operational risk while maintaining the high standards of service delivery citizens expect. By providing realistic test data on demand, organisations can shorten test cycles, increase deployment frequency, and confidently deploy services that protect citizens and have been validated against real-world scenarios and edge cases. Teams regain confidence and capability as data access barriers are removed through modern practices that protect citizen privacy.

2i's Synthetic Test Data Generation service ensures digital services are tested thoroughly when it matters most with realistic data, at realistic scale, complete confidence in data protection compliance, and without compromising the privacy of the citizens your services support.

2 Benefits

- Deploy digital services with confidence in thorough, realistic testing
- Protect organisational reputation by meeting data protection requirements by design
- Enable faster, more frequent delivery by removing test data bottlenecks
- Test edge cases and rare scenarios that strengthen service quality
- Scale test data volumes for performance and load testing requirements
- Meet GDPR, DPA 2018, and sector-specific data protection commitments
- Reduce project delays with on-demand, compliant test data provisioning
- Increase stakeholder confidence with auditable data protection evidence

3 Features

- Partnership with Tonic (industry-leading synthetic data platform) plus 2i proprietary tooling
- Statistically realistic data generation maintaining production distributions and patterns
- GDPR-compliant by design with no production data or PII in test environments
- Edge case generation for scenarios difficult to find or extract from production
- Referential integrity preservation across complex data relationships and tables
- On-demand data provisioning integrated into test environment workflows
- Scalable generation from hundreds to millions of records as testing requires
- Audit-ready documentation providing compliance evidence for regulators

4 Deliverables

Our deliverables provide the evidence and documentation you need:

- Test Data Requirements Analysis
- Data Model and Relationship Mapping

- Synthetic Data Generation Strategy
- Data Generation Model Configuration
- Generated Test Datasets (validated against requirements)
- Edge Case Scenario Specifications
- Statistical Validation Report
- Referential Integrity Verification
- Compliance Documentation and Audit Trail
- Test Environment Integration Guide

5 How it Works

To deliver consistent, high-quality synthetic test data, we take a proactive, structured approach with strong stakeholder engagement, underpinned by deep data privacy and technical delivery expertise. This ensures all relevant information such as technical, operational, compliance, and organisational factors is captured, analysed, and translated into a test data approach that all stakeholders can understand and use to make evidence-based decisions with certainty and confidence.

5.1 Engage – Establishing trust and shared intent

Effective test data solutions start with strong relationships and a thorough understanding of the data landscape. Before defining generation, strategies or building data models, we invest time in understanding the people behind the testing need and the data itself: delivery teams, data owners, compliance stakeholders, and decision-makers.

We work with stakeholders to answer the critical questions that underpin our Analysis stage: Where is the data? Is it suitable for synthesis? What logic is embedded in the data? Through early conversations and structured assessment, we map data sources, identify entity relationships, understand business rules and constraints, and evaluate the complexity of the data landscape.

This early engagement ensures synthetic data generation is aligned to real testing challenges, citizen privacy protection, and shared objectives, forming the foundation for effective collaboration throughout the engagement.

5.2 Discover – Understanding the service and its risks

Effective test data solutions start with strong relationships and a thorough understanding of the data landscape. Before defining generation, strategies or building data models, we invest time in understanding the people behind the testing need and the data itself: delivery teams, data owners, compliance stakeholders, and decision-makers.

We work with stakeholders to answer the critical questions that underpin our **Analysis** stage: Where is the data? Is it suitable for synthesis? What logic is embedded in the data? Through early conversations and structured assessment, we map data sources, identify entity relationships, understand business rules and constraints, and evaluate the complexity of the data landscape.

This early engagement ensures synthetic data generation is aligned to real testing challenges and shared objectives, forming the foundation for effective collaboration throughout the engagement.

5.3 Solution

Based on shared understanding and insight, we work side by side with clients to co-create a tailored synthetic data solution with Privacy embedded by design.

Rather than imposing predefined models, we collaboratively define:

- Scope and aims aligned to testing requirements
- Data volumes, complexity, and edge case coverage needed
- A generation strategy using Tonic, 2i tooling, or both
- Sensitivity status assigned to every data element
- Suitable generators identified for each sensitive field
- A comprehensive privacy scan across the dataset, producing a detailed Privacy Report

The Privacy Report provides auditable evidence that all sensitive data elements have been identified and that appropriate generation strategies are in place—meeting GDPR, DPA 2018, and sector-specific data protection requirements. This collaborative approach builds ownership and ensures the solution delivers practical, usable, and compliant test data.

5.4 Mobilise – Turning strategy into action

Once the approach is agreed and privacy controls are in place, we configure and execute the data generation pipeline. This is our ****Synthesis**** stage: choosing the correct data generators for each data element, configuring the pipeline to maintain statistical distributions, referential integrity, and business rule compliance, and running the pipeline to produce synthetic datasets at the required volume and complexity.

Edge case scenarios are deliberately generated to ensure testing covers the conditions that matter most, including rare but critical scenarios that are difficult to find in production data. We establish clear communication channels, outcome-focused reporting, and proactive risk management to ensure transparency and confidence throughout execution.

As stakeholders have been involved in shaping the solution, teams move forward with clarity and ownership. This clearly defined, adaptable mobilisation approach transforms uncertainty into confidence and ensures synthetic data generation is integrated smoothly into ongoing testing and delivery workflows.

5.5 Deliver – Evidence, assurance, and lasting confidence

Delivery focuses on measurable outcomes and meaningful impact.

Before data reaches test environments, we run our Evaluation stage: comparing original and generated data using our evaluation scripts, assessing statistical fidelity, referential integrity, distribution accuracy, and edge case coverage, and producing a detailed Quality Report that provides evidence the synthetic data meets the defined requirements and is fit for purpose.

Validated synthetic data is then delivered through the method that best fits the organisation's workflows—via the Platform for on-demand provisioning or integrated directly into the CI/CD pipeline via API for automated test environment setup.

Outcomes typically include:

- Reduced risk of quality issues affecting critical public services through comprehensive testing with realistic data
- Clear evidence of thorough testing with realistic, compliant data
- Early identification and resolution of edge case scenarios
- Increased confidence in go/no-go decisions based on comprehensive testing
- Improved delivery velocity through on-demand data provisioning
- Audit-ready documentation demonstrating data protection compliance, including Privacy Reports and Quality Reports

Beyond delivery, we leave behind increased capability, clarity, and confidence. The measure of success is not just generated datasets, but the certainty and assurance organisations gain to test and deploy digital services with complete confidence in data protection compliance.

6 Pricing

Services are priced based on client requirements and delivery model. Engagements can be delivered on a **Time & Materials** or **Fixed Price** basis, depending on scope, risk profile, and complexity.

Pricing is designed to demonstrate value for money through risk reduction, delivery efficiency, and improved service outcomes. Please contact us for any further information.

7 Why 2i

2i is a multi-award winning, independent QA and testing consultancy headquartered in Edinburgh, with UKwide teams and capacity in Coimbatore, India. Since 2005, we've delivered confidence through assurance, reducing risk, accelerating delivery, and embedding quality with strategic test consultancy, AI-enabled test automation, and lean delivery alongside client and multi-supplier teams.

In late 2024, we strengthened our position by acquiring nFocus, combining two leading consultancies into one of the UK's largest testing teams: 400+ professionals, including 240+ focused on test automation. This scale enables rapid mobilisation across major UK locations and flexible resourcing as needs change.

We operate across highly regulated sectors and major public sector programmes, leading on quality engineering and AI assurance where required. Clients gain a trusted partner that integrates quickly, upskills in-house teams, and delivers predictable outcomes, strengthening quality and operational confidence.

We're people-first and Great Place to Work® accredited for three consecutive years. As a Scottish based company, we also commit to community benefits, Fair Work First practices, above-living-wage pay, 4% revenue reinvested in learning, and a carbon reduction plan supporting net-zero goals.

8 Client Value

We currently deliver testing services to several clients in the Scottish Government and UK Central Government, including on critical national infrastructure projects where we have de-risked some of the largest and most complex delivery programmes in Europe.

During our partnership with the Scottish Government, we won an award at the European Software Testing Awards in 2021.

9 Contact

Call +44 (0) 131 220 6888 or email info@2itesting.com to discuss your technology challenge and learn how 2i can help you.