

expo IQA 25

MADRID
May 20th,
21st & 22nd
2025



expoqqa.eu

¡BIENVENIDOS!

Real World Data vs Fake Data: Finding the Best Fit for Testing



Ana Victoria Rodríguez

Product Architect

[Ana Victoria Rodriguez | LinkedIn](#)

**Automatizar pruebas
no sirve de nada si
los datos fallan.**

Los **equipos de QA** automatizan pruebas, pero se estancan siempre en el mismo punto: **los datos de prueba.**

Sin **datos realistas, accesibles y disponibles** no hay automatización que valga.

Casos de prueba con TDM



Diseño del caso de prueba

Diseño del caso de prueba, los **datos de entrada** y los resultados esperados.



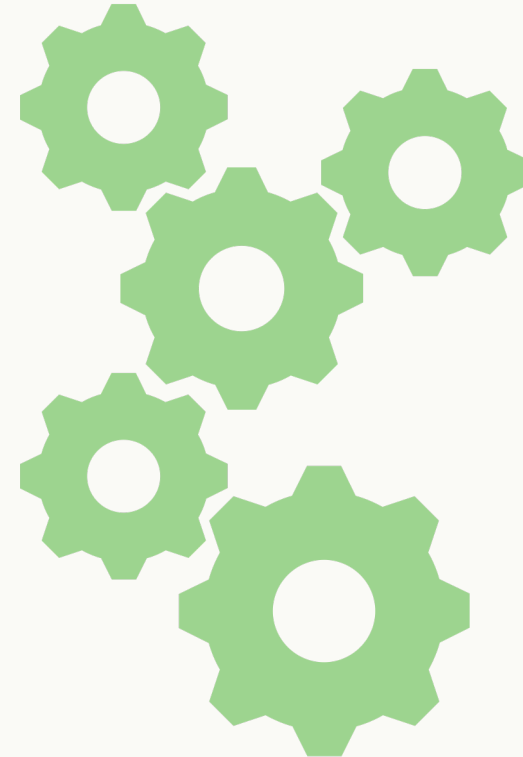
Ejecución del caso de prueba

Inyección de los datos de entrada y ejecución.



Verificación de resultados

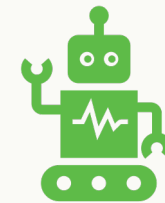
Verificación del resultado.



Real
Data



Fake
Data



¿Qué son los Real World Data o Datos del Mundo Real?

Datos derivados de producción, anonimizados pero que conservan el valor técnico y contextual necesario para pruebas realistas.

- ✓ Capturan **comportamientos y lógica de negocio reales**, incluyendo patrones complejos, fluctuaciones y sesgos.
- ✓ Reflejan la **evolución** real del dato en el tiempo.
- ✓ Mantienen **integridad referencial**.
- ✓ Conservan la **distribución** estadística auténtica.
- ✓ Permiten identificar errores reales: nulos, datos inconsistentes, casos límite.
- ✓ Soportan pruebas **end-to-end** e integración con otros sistemas.

Los **datos reales** son **insustituibles** por su valor técnico, pero sin una anonimización adecuada y control operativo, pueden convertirse en un riesgo más que en una solución.

TDM

Fake Data o Datos Sintéticos

El dato sintético o fake data, es un término ampliamente utilizado en la industria, pero **carece de una definición** clara y universal.

Las interpretaciones más comunes incluyen:

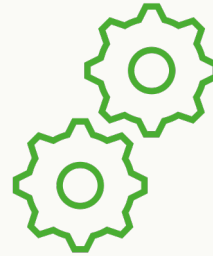
1. **Datos generados manualmente (con hojas de cálculo, scripts, API's)**
Rápidos, pero frágiles, costosos y con poca cobertura.
2. **Datos generados por IA (entrenados con patrones reales)**
Potencialmente realistas, pero hoy en día no son fiables para entornos críticos.

Soluciones habituales pero ineficientes



Hojas de cálculo

¿Cuánta complejidad estructural de datos puedo representar en una hoja de cálculo?



APIs de negocio

¿Cuántos procesos de negocio tengo cubierto con APIs? ¿Quién desarrolla las que faltan? ¿Quién más necesita mis APIs?



Datos reservados

¿Cuánto trabajo manual dedico a crear datos y coordinar su uso? ¿Con qué resultado?



Scripts de automatización

¿Cuánto me cuesta que los ingenieros de automatización desarrollen y mantengan scripts? ¿Cuánta casuística cubren?

El dato sintético es insuficiente

- ✗ Sin Antigüedad
- ✗ Falta de complejidad
- ✗ Ausencia de escenarios poco frecuentes
- ✗ Demasiado perfecto
- ✗ Volumen insuficiente
- ✗ Distribución estadística no representativa

El problema no es el dato sintético, sino cómo se genera

Un **enfoque** basado en el modelo de datos genera **datos sintéticos con valor real**.

- ✓ Mantienen coherencia, estructuras y reglas de negocio **reales**.
- ✓ Reproducen la **complejidad** y la lógica de los entornos productivos.
- ✓ Conservan integridad referencial y distribución estadística realista.
- ✓ Incorporan **diversidad**, volumen y **casuística** relevante.
- ✓ Están libres de información sensible real, **cumpliendo** con **normativas** de privacidad.
- ✓ Se **adaptan** a necesidades específicas: nuevos modelos, pruebas de carga, escenarios extraordinarios.

Flujo de la generación de datos



No se trata de elegir entre **datos reales** o **datos sintéticos**, sino de **aplicar el tipo de dato más adecuado** según el objetivo de cada prueba.



Una estrategia efectiva de testing combina ambos para maximizar cobertura, realismo y cumplimiento.



¿Cuándo usar cada tipo de dato?

Los datos del mundo real son la mejor opción debido a su riqueza. Sin embargo, los datos sintéticos son especialmente útiles en situaciones donde los datos reales no existen o no son suficientes.

Nuevos desarrollos

Cuando se modifica el modelo de datos y aún no hay registros disponibles en los entornos existentes.

Casos poco frecuentes

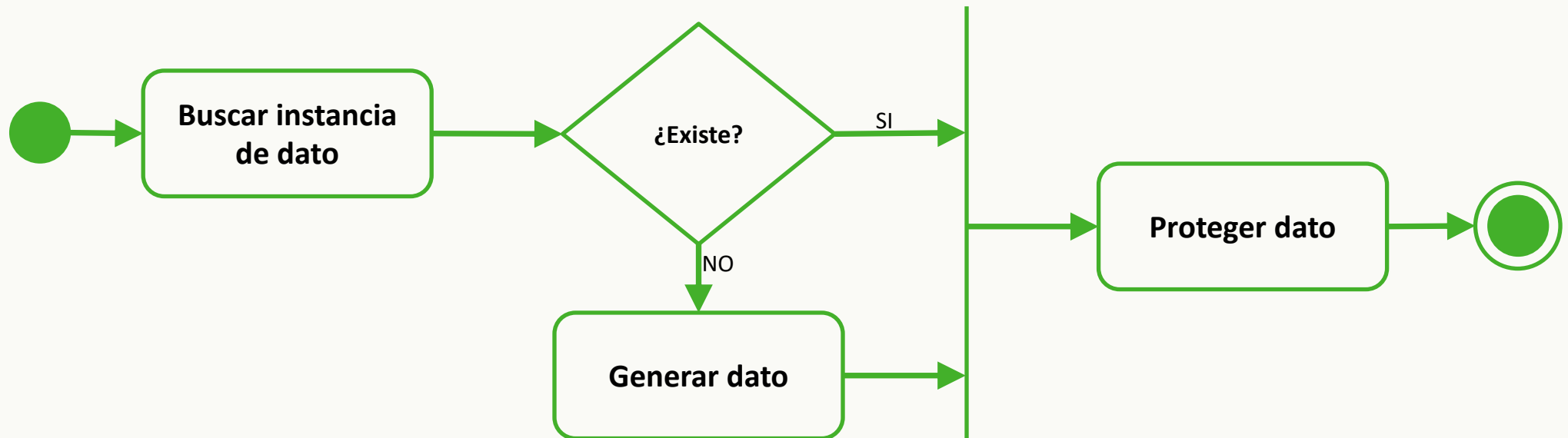
Cuando se necesitan múltiples instancias de un escenario raro que solo aparece una o dos veces en la realidad.

Datos de entrada con TDM



Diseño del caso de prueba

Diseño del caso de prueba, los **datos de entrada** y los resultados esperados.





01 REQUEST TEST CASE INFORMATION

02 CREATE TEST CASE DATA SESSION

Test Automation Orchestrator Tool

05 ORDER TEST CASE EXECUTION

Test Case Management Tool

Test Case Execution Tool

leapwork

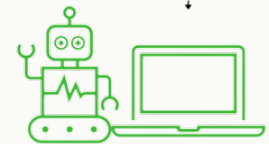
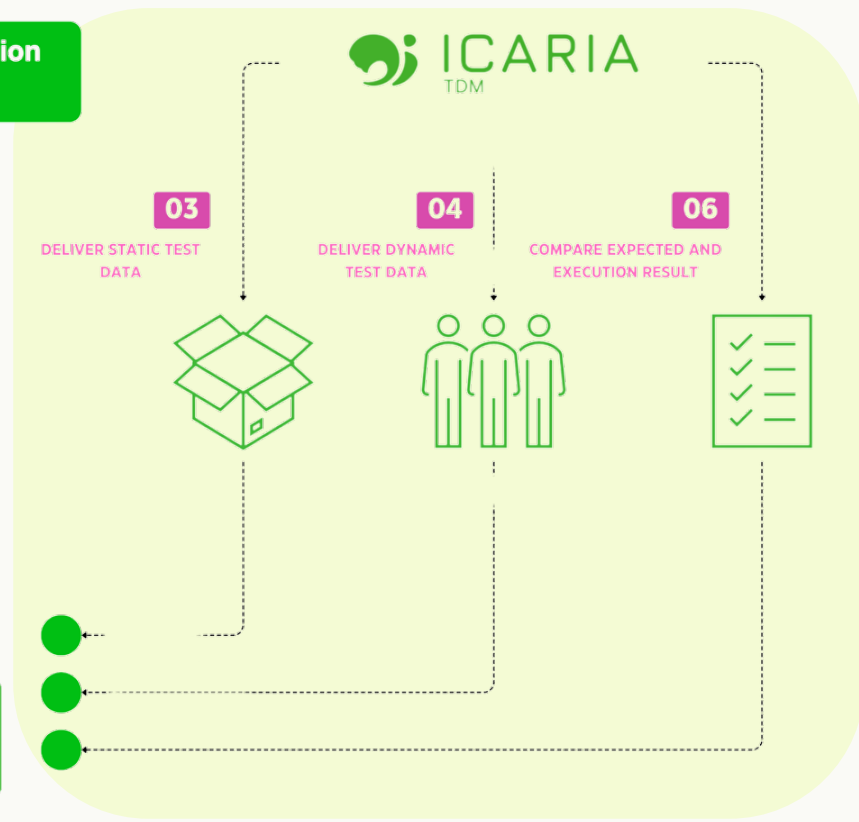
Selenium

Playwright

TestLink

TESTDIRECTOR

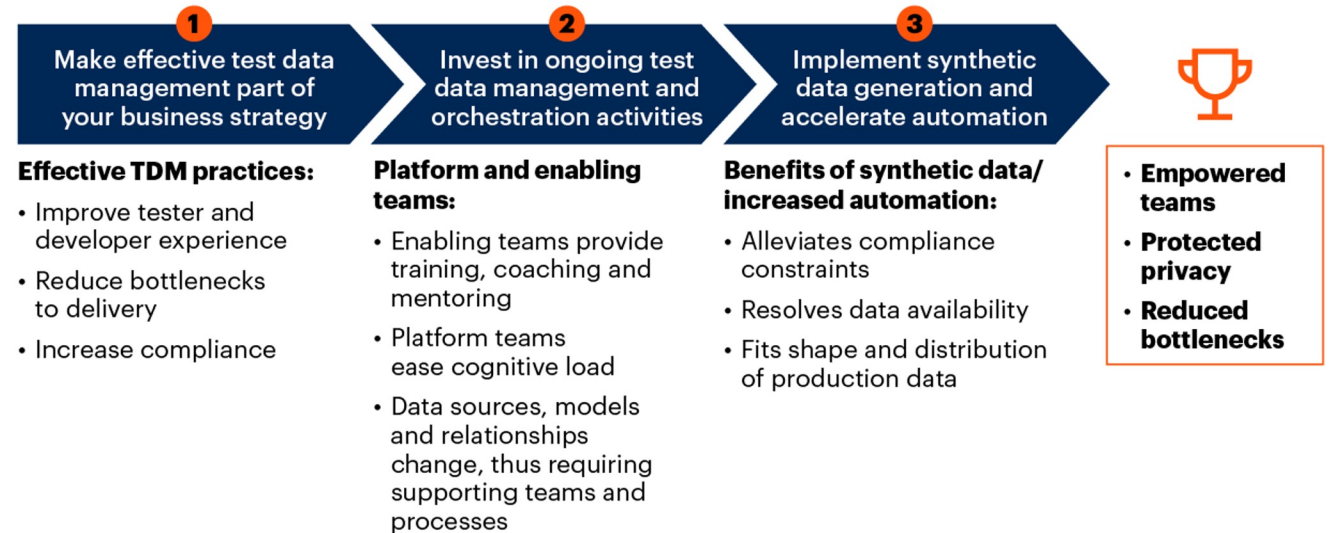
Jira Software



Según Gartner, la adopción de buenas prácticas en TDM es incipiente y desigual.

Es por ello que se sugiere integrar la estrategia de TDM a la estrategia del negocio, así como invertir en plataformas.

Three Steps for Test Data Management



Source: Gartner
TDM: Test data management
818798_C

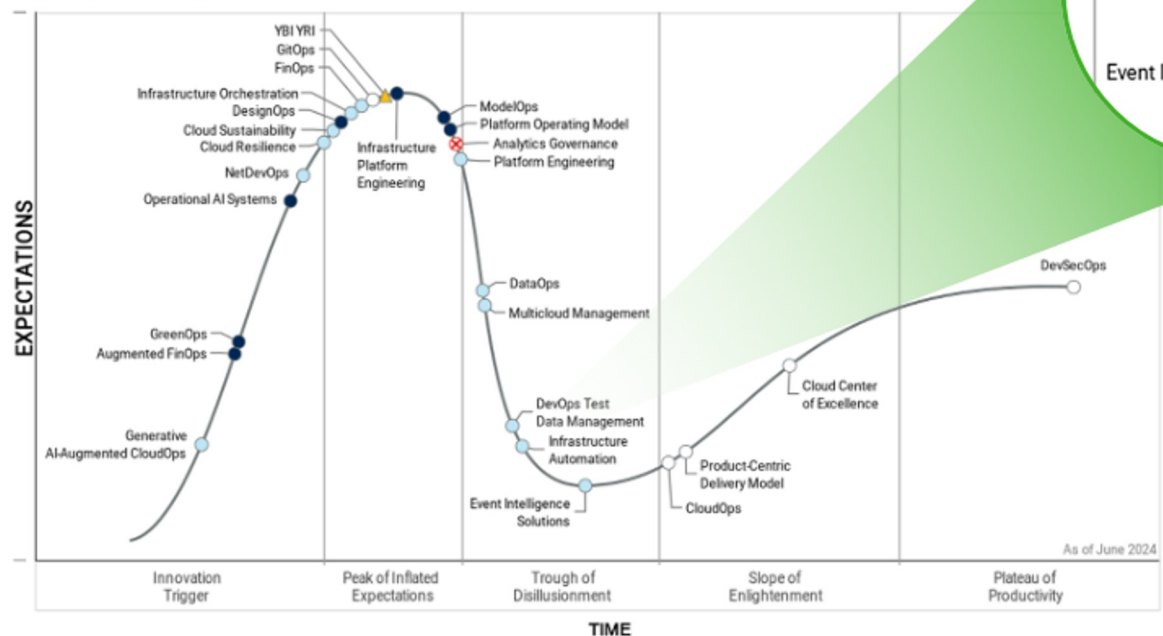
Gartner

La tecnología ha madurado.

La categoría de TDM ha superado el pico de la sobreexpectación, ofreciendo resultados fiables.

Figure 1: Hype Cycle for Operation Models, 2024

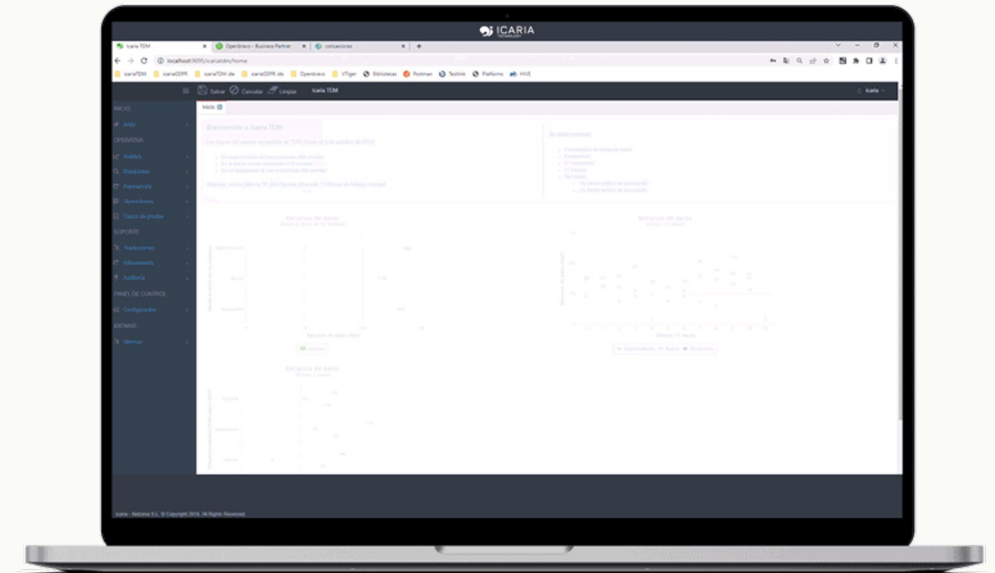
Hype Cycle for Operation Models, 2024



Gartner

icaria TDM es la plataforma de Test Data Management diseñada para aplicaciones críticas y OSS/BSS.

Ofrece **datos realistas, seguros, correctos y coherentes**, exactamente cuando y cuantas veces se necesiten. Garantiza la **fiabilidad de las pruebas**, incluso en los casos más complejos, sin comprometer la agilidad.



Con **icaria TDM**, tus pruebas contarán con:

- ✓ **Más cobertura** → Asegura que cada prueba refleje la complejidad del mundo real, incluso en escenarios difíciles de replicar.
- ✓ **Más velocidad** → Minimiza los tiempos de espera y automatiza la provisión de datos, acelerando el desarrollo y la entrega.
- ✓ **Menos riesgos** → Cumple con GDPR y otras normativas, garantizando la seguridad de los datos en cada etapa del testing.
- ✓ **Más eficiencia** → No sólo acelera los ciclos de prueba, también reduce los costes que habitualmente pasan desapercibidos, alcanzando un ROI del 309%.

¿Quieres evaluar tu proceso actual de TDM?

Podemos ayudarte a identificar cuellos de botella y costes ocultos.

Reserva una reunión 1:1



Q&A:
Resolvemos vuestras
consultas.

¡MUCHAS GRACIAS!

Real World Data vs Fake Data: Finding the Best Fit for Testing



Ana Victoria Rodríguez

Product Architect

[Ana Victoria Rodriguez | LinkedIn](#)

expo **QQA** 25

MADRID
May 20th,
21st & 22nd
2025

Thank you for attending

expoqqa.eu