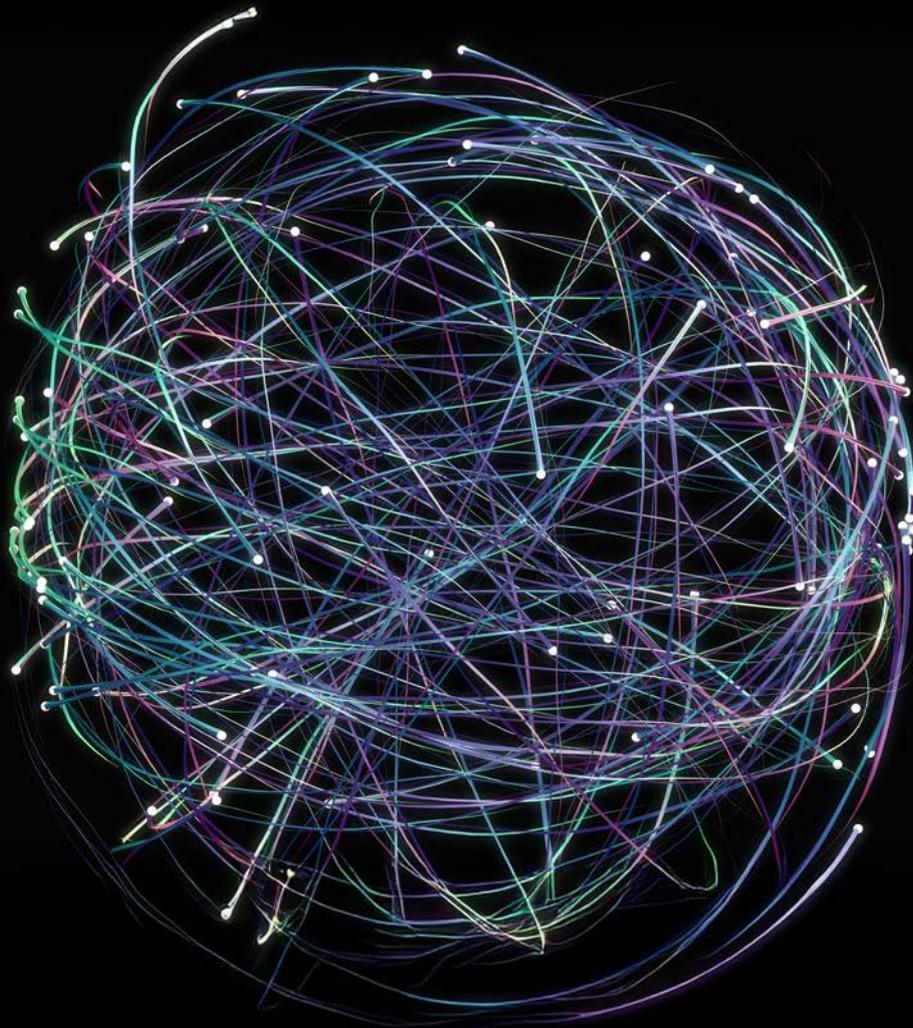# FCA Synthetic Data Anti-Money Laundering (AML) Solution Sprint

13 May - 8 July 2026

"The Synthetic Data AML Solution Sprint represents a bold step forward in our commitment to innovation and fight against financial crime. By making realistic synthetic data available, we're not only accelerating the development of cutting-edge technology across the ecosystem - from transaction monitoring solution providers to firms building their own AML capabilities - but also reinforcing the UK's leadership in the global fight against financial crime. This initiative is about empowering innovation that delivers real-world impact for consumers."

Steve Smart, Joint Executive Director of Enforcement and Market Oversight

Contents

Introduction

Financial institutions continue to face formidable challenges in Anti-Money Laundering (AML) due to the intricate nature of illicit financial flows, a rapidly evolving regulatory landscape, operational constraints, data and technology challenges and more.

Despite actions being taken across industry, significant issues remain. Current UN estimates indicate that between 2% and 5% of Global GDP - equivalent to approximately $800 million – $2 trillion[1] - is laundered annually.

Transaction Monitoring is key to tackling money laundering by identifying customers acting in an unusual or suspicious manner. Firms, however, have highlighted to us that they are often unable to access suitable transactional data to develop and test their solutions due to stringent privacy regulations that restrict sharing of this highly sensitive data. By developing realistic synthetic data sets, we can enable firms to develop and test their models to better detect money laundering whilst at the same time meeting these privacy requirements.

What we have done so far

Recognising the potential for this technology, the FCA set up the Synthetic Data Expert Group (SDEG) in March 2023. The group brings together 21 experts from across the financial services industry, public sector, data and technology vendors, and consumer groups to further explore the use of synthetic data in financial markets (e.g. data augmentation, testing and model validation, internal and external data sharing).

In September 2023, we launched the Authorised Push Payment (APP) Fraud Synthetic Dataset in the Digital Sandbox, which has successfully supported 65 users across 42 projects.

Aims and outcomes

As part of this wider programme on Synthetic Data we have been exploring how we can generate and use Synthetic Data across a range of different use cases to support the fight against financial crime. This aligns with **FCA's recently refreshed Strategy**, which prioritises actions against financial crime.

---

1   www.unodc.org/unodc/en/money-laundering/overview.html

We have built a new Synthetic Data set with the help of our delivery partners, The Alan Turing Institute, Plenitude Consulting and Napier AI, which uses real transactional data, provided with consent from UK Retail Banking to generate a fully synthetic transactional payments data set. This data set has then been augmented with synthetic money laundering typologies - based on data collected from documented real world money laundering examples.

The aim of this Data Sprint is to make this data set available through the **FCA Digital Sandbox** to firms who are developing Transaction Monitoring solutions - with a particular focus on new and emerging techniques (e.g. AI). We want firms to use this data over the course of the sprint, and come back together to share findings on how synthetic data can be used to foster innovation in the detection of money laundering.

Our goal is to support firms and to provide a platform to understand how:

1. New technologies can help in the detection of financial crime
2. Synthetic data can be used to effectively test and develop transaction monitoring solutions

We hope that through this exercise we can offer a platform for solution providers to demonstrate how new technologies can be used to help accelerate innovation in anti-money laundering technology development. Whether testing novel approaches or enhancing established systems, participants will have the opportunity to explore the potential of synthetic data in advancing AML capabilities.

Event and participation details

When
13 May until 8 July 2026

-

What
The Data Sprint is an 8-week cohort-based event to support firms who are developing Transaction Monitoring solutions to detect and prevent money laundering, with a particular focus on new and emerging technologies such as AI, allowing them to explore the data sets and demonstrate their tools using the FCA AML synthetic datasets.

We also are extremely keen to get feedback on the data sets from the teams who have been using them - to iterate and enhance the data sets before potentially making them available to a wider audience.

Throughout the course of the Data Sprint, participating teams will have access to dedicated FCA coordinators who will provide relevant support where appropriate.

-

Where
13 May: Virtual kick off and Bootcamp Briefing.
Virtual weekly engagements with the FCA throughout May and June.

8 July: In person Demo Day at the FCA Stratford office, 12 Endeavour Square, London, E20 1JN.

-

How to apply
Applications open on 24 March 2026 and close on 26 April 2026.
Application results will be communicated by 6 May 2026.

Applications will be accepted via the **FCA Innovation Platform**.

Problem statements and guidance

There are two problems statements, firms can choose one of these problem statements to work on over the course of the Data Sprint.

Platform for demonstration of innovation in Transaction Monitoring

To understand whether synthetic data be used successfully by firms to demonstrate the effectiveness of new and emerging technologies (e.g. AI) in the fight against Money Laundering.

The data set will be provided **without** Money Laundering labels (i.e. the transactions that relate to money laundering typologies are not marked within the data set), and firms will be able to use the data set to demonstrate how their new technologies can detect the money laundering typologies within the data.

Key areas of focus:

- Can the synthetic data be used to demonstrate capabilities in performance, explainability, efficiency and accuracy?
- Does the synthetic dataset enable experimentation with techniques that would be difficult to test on real data?

Enhance existing Transaction Monitoring solutions

To understand whether firms are able to leverage our synthetic dataset to enhance their existing money laundering detection systems, particularly through the application of new and emerging technologies (e.g. AI).

The data set will be provided **with** Money Laundering labels (i.e. the transactions that relate to money laundering typologies are marked within the data set), and firms will be able to use the data set to test and train their existing solutions.

Key areas of focus:

- Can the dataset be used to help improve existing solutions by providing alternative examples of money laundering?
- Does the AML synthetic dataset reflect realistic transaction patterns, typologies, and behaviours when compared to data previously observed?

The data

Data Sprint participants will be able to access the following FCA datasets (hereinafter, the "Synthetic AML dataset"):

- AML Data Sprint - Synthetic current account transactions
- AML Data Sprint - Synthetic bank accounts details
- AML Data Sprint - Synthetic customers details
- AML Data Sprint - Synthetic branches details

The synthetic AML dataset is not identical to the original data collected from UK retail banking. Instead, we have taken the submitted data and developed a synthetic dataset that preserves the relationships across the tables. We then augment this dataset with money laundering transactions, informed by our research into common typologies and laundering techniques, resulting in the final synthetic dataset.

Participants will access data via the Sandpit, a secure coding environment within the **FCA Innovation Platform**. If needed, data can also be exported for analysis off platform during the Data Sprint, subject to signing an additional Non-Disclosure Agreement (NDA).

The inclusion of money laundering labels in the synthetic current account transactions depends on the problem statement you choose:

- **Problem Statement 1**: Labels are not included.
- **Problem Statement 2**: Labels are included.

We will provide detailed instructions in the Bootcamp Briefing, along with data schemas and dictionaries for each of the datasets listed above. Participants will be given access to a standard cloud-based workspace hosted on Amazon Web Services, with specific details shared nearer to the Data Sprint. At first, participants' cloud-based workspace will have internet access to install necessary packages. After this initial setup period, the internet will be disabled, and the datasets will be loaded directly to participants' workspace. Alternatively, the data can be removed from the platform (subject to security approval) and the analysis performed off platform.

The data sets that will be provided have been generated using cutting edge synthetic data generation approaches with a focus on reflecting real world conditions. We would like to use this sprint as a collaborative exploration of the data and its utility in demonstrating the effectiveness of innovation in Transaction Monitoring. As you explore the data and use it throughout the sprint we are very interested in exploring with you any updates or enhancements that we can make to the data so that we can iterate and improve it for future use cases.

Building the teams

To support the best outcomes for the Data Sprint, we encourage participation from a diverse and wide range of participants who are already in the process of developing solutions to manage and mitigate money laundering risks.

For the purposes of this Data Sprint, we ask participants to apply as a fully formed team, with a minimum of 2 team members and a maximum of 6 members.

Teams would benefit to be made up of the following attributes:
Note, one person can take on multiple roles and you are not expected to fulfil each and every one of the following roles - this is an indicative guide to help build your team:

- **The Visionary** (Innovative Ideas Person)
  You bring a seed of an idea and build a team around it. You are creative and have a flair for innovation. You know how to shape the design to make the moving parts fit together seamlessly.

- **The Juggler** (SCRUM Master)
  Every team needs a juggler! You will help ensure the smooth running of the event for your team. You will unblock the blockers, facilitate collaboration within the team and work with your FCA central coordinator and the FCA TechSprint team to feedback progress and highlight any issues.

- **The Data Specialist** (Data Experts)
  As a central player in the team, you are able to work well with data to be able to find patterns and are prepared to manipulate it for good. You feel comfortable working with the datasets provided. You're able to tell a story with numbers and data, and can bring information together to form a compelling narrative.

- **The Closer** (Marketing Guru)
  You know how to make ideas shine. You are confident speaking in front of others and can explain your team's solution clearly, engaging, and persuasive. You help shape the story behind the data testing and model development and make sure it connects with the audience.

Building your application

1. Complete the [registration form](#) by 26 April 2026, giving as much information as possible about relevant experience and initiatives in this space.

2. Provide an overview of the problem statement you will address, and how your proposed data testing plan aligns with the purpose of the Sprint.

3. Share any additional resources and expertise you may require during the sprint so we can match you with the relevant experts.

4. Confirm you are able to commit to the Data Sprint Virtual Bootcamp Briefing on 13 May 2026 and the in-person Demo Day on 8 July 2026 at the FCA's London Office.

5. Let us know if you have any additional requirements or adjustments to help you take part. FCA Sprints are inclusive events and we aim to accommodate requests.

To give yourself the best chance of being selected, below are a few guidelines:

- Make sure your application is clear in how it addresses the intended problem statement

- Provide details of the value you can add and be clear on the main focal point for your testing plan and solution

- Provide a high-level overview of the potential risks of the testing plan and solution and explain how those risks could be mitigated

- Provide details of any examples or case studies, either from personal experience or through your research, that demonstrates the value of your solution

- Summarise how your solution is unique in addressing the challenges set out in our problem statement

Following the closure of the application period we will:

- Review the applications and inform you on whether your application has been successful by 6 May 2026

- Provide full onboarding instructions for the Data Sprint kick off and Bootcamp Briefing event on 13 May 2026

The spirit of the event

Collaboration
The AML Synthetic Data Sprint is an open event, bringing together a range of industry participants and stakeholders. We expect participants to be open to sharing findings with other participants.

Balanced participation
We aim to provide opportunities for a diverse group of entities to participate in the event and encourage participation across various industries e.g., tech firms, RegTechs, academia, consultancies, think tanks etc.

Logistics
Participants will be required to self-fund participation in the event, including any travel and accommodation. Catering will be provided for the in-person Demo Day.

Intellectual property
The AML Synthetic Data Sprint is an ideation event, and we ask that you participate as collaborative inventors. Participation means engaging in the spirit of the event by showcasing the analysis undertaken at the end of the event. You retain all IP rights in your existing solutions and any outputs developed through participation in the Sprint.

Communications approach and guidelines

We are supportive of participants discussing their involvement in our AML Synthetic Data Sprint.

Once participants are confirmed by us by early May, participants can share supporting communications via social media, as well as during and after the AML Synthetic Data Sprint using *#FCASprint #AMLSyntheticDataSprint*.

Press
Any wider communications (including digital content or press releases) that refer to the AML Synthetic Data Sprint must be agreed with the FCA Press Office prior to release.

Nature of participation
Participants are advised to use the term "collaboration" instead of "partnership" in any communications. For any further queries, please contact the FCA TechSprint Team at techsprints@fca.org.uk.