CITY OF LONDON

FINANCIAL CONDUCT AUTHORITY

Supporting innovation in ESG data and disclosures: the digital sandbox sustainability pilot

June 2022

Contents

Moving around this document

Use your browser's bookmarks and tools to navigate.

To **search** on a PC use Ctrl+F or Command+F on MACs.

1 Executive Summary

1.1 One of the FCA's three objectives is to promote competition in the interest of consumers. An essential component of effective competition, and one of its key drivers, is innovation.

1.2 Creating a regulatory environment that supports responsible innovation ultimately benefits consumers and firms. It both provides inventive solutions to consumers' needs and enables agile start-ups to challenge incumbents. This in turn drives incumbents to compete harder to retain customers.

1.3 Since its inception in 2014, the FCA's Innovation division has evolved a suite of market-facing services to support innovative firms and the competitiveness of UK markets. These services include TechSprints, the Regulatory Sandbox, and Innovation Pathways, which support firms at multiple levels of maturity, from initial ideation and proof-of-concept (PoC), to obtaining authorisation, and scaling internationally through cross-border testing.

1.4 In partnership with the City of London Corporation (CoLC), in 2020 we piloted the 'Digital Sandbox' to support firms at the 'proof of concept' stage of development. The Digital Sandbox provides participants with a suite of features including bespoke synthetic data assets and a range of relevant mentors, to develop solutions from both technological and conceptual angles. The pilot focused on challenges exacerbated by the Covid-19 pandemic.

1.5 Since the pilot concluded, the Kalifa Review of UK FinTech has recommended that regulators create a permanent digital sandbox to encourage further collaboration within UK financial services. In response, we have sought to iterate and improve the digital sandbox model by running a second pilot, focused on solving challenges for ESG data and disclosures.

1.6 Following the successful conclusion of the second pilot, we are committed to building on the lessons learned during the previous 2 years to inform a future and permanent operating model for the digital sandbox.

1.7 In this report, we provide an overview of the digital sandbox sustainability pilot and discuss the lessons learned. It also outlines our current thinking for the future of the sandbox. We hope this report will contribute to the current dialogue surrounding innovation, as well as be of interest to regulatory peers, as we look to progress future iterations of the digital sandbox.

Key lessons

1.8 Based on participants' feedback and our experience of hosting the second pilot, we have identified several key lessons:

- Participants unanimously agreed that the digital sandbox initiative has a significant impact on innovation acceleration, citing networking and collaboration features as

being of key value. Other benefits include product testing and validation, access to data assets, and the opportunity to demonstrate solutions to an audience of industry experts via showcase days.

- Participants and mentors praised the sustainability pilot for bridging the gap between financial services and sustainability. However, ESG data and disclosures is also a broad theme that can sometimes be difficult to link with financial services. Due to the breadth of the field, the quality of the synthetic data assets and the ability to meet all mentor requests from participants sometimes suffered.
- ESG reporting and disclosures is a relatively new field in which open and rich datasets are not always readily accessible. In some cases, data did not yet exist to meet participants' requirements or could not be procured within acceptable timeframes. When selecting focus areas for future digital sandbox initiatives, due consideration should be made of the data available in the field to ensure that there is enough quality data to meet participants' testing requirements.
- The process of tailoring synthetic datasets required longer timescales than the approach used for the first pilot, resulting in some data assets not being available at the start of the sustainability pilot. Future iterations of the sandbox should consider how to balance producing bespoke data sets in the timescales available, and whether it is more appropriate to generate generalised synthetic data assets (not specifically tailored to participants' requirements) before the pilot starts.
- The teams experienced challenges in collaborating with other teams, particularly as the pilot progressed. A more structured approach towards facilitating collaboration between different teams may be helpful for future iterations of the digital sandbox. This could include more in-person activities and events where possible.
- Continued access to the digital sandbox's platform and services, particularly the data assets, would generate significant benefits for participants. We are currently exploring options to enable long-term access to the digital sandbox data for alumni, as well as opening up data access to a wider audience in a GDPR-compliant way.
- Both the usage statistics and participants' feedback suggest that, while some features of the platform provided value for users, the platform itself was not widely used over the course of the pilot. Users tended to extract relevant information from the platform, such as contact details for mentors or data assets, and use this information off-platform. We have identified several ways to refine the data and networking features that could improve the overall user experience.

Next steps

1.9 The Digital Sandbox Pilot and Sustainability Pilot has demonstrated that collaboration and access to data can stimulate beneficial innovation in the market. We are committed to establishing a permanent operating model for a digital testing environment based on the principles, and improved by the experience and feedback, of the Digital Sandbox initiative. We are currently undertaking further research and industry engagement to ensure that this future service meets the needs of its end-users, and to understand the most relevant potential use cases that this initiative could support.

1.10 Beyond the application of a digital testing environment and collaboration platform to support market innovation and our existing regulatory services such as the Regulatory Sandbox and TechSprint programme, we are also exploring further benefits of the platform, including its ability to facilitate regulatory-to-regulatory collaboration in, for example, the development of SupTech (supervisory technology) solutions.

1.11 We are also exploring methods to make the digital sandbox data assets openly accessible in a way that complies with data protection laws. Datasets would be available as part of a public good designed to promote innovation and competition across financial services.

2 Digital sandbox sustainability pilot overview

2.1 ESG data and disclosure is a corporate priority for both the FCA and the City of London Corporation. As outlined in our Business Plan for 2022/23, the FCA is committed to promoting positive change by delivering on our recent ESG commitments. It is a key area of interest and one that we need to tackle together in co-operation with the industry.

2.2 The FCA and CoLC ('the Working Group') hosted the digital sandbox sustainability pilot between November 2021 and March 2022.

Use cases

2.3 In conjunction with a broad range of industry stakeholders and internal colleagues, we hosted several workshops to gather problem statements, hypothesize solutions, and identify data inputs. Following the workshops we published a report which outlines in more detail the process for identifying use cases for this pilot.

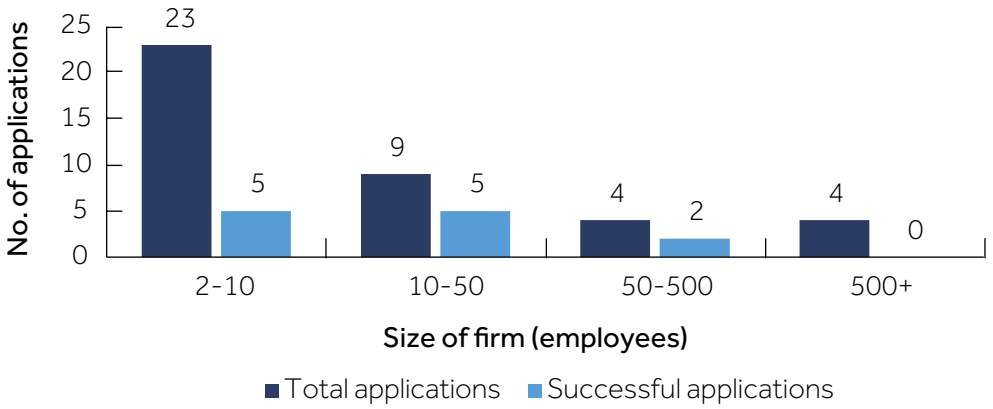
2.4 Applicants were asked to solve one of the following use cases:

- **UC1 – Improving Consumer Understanding:** How can technology help consumers understand the ESG characteristics of the products and providers they engage with, as well as provide visibility around alternatives aligned with their needs and preferences?
- **UC2 – Transparency and Disclosure:** How can technology enable transparency in disclosure and reporting on sustainability, especially on the characteristics of corporate assets and the profile of their supply chains?
- **UC3 – Validation of ESG Data:** How can technology be used to automate the assurance of a listed issuer's ESG data and validation of its ESG-labelled corporate bond issuance? (e.g. IoT, decentralised ledgers, centralised platforms, satellite imaging, AI)

Application snapshot

2.5 We received applications from a range of firms of varying sizes, with the most common applicants being small firms/start-ups. Mid-size firms (10-50 and 50-500 employees) had significantly higher application success rates of 50%+, whereas smaller firms (2-10) had a 22% success rate. The quality of application from start-ups varied, with many applicants from this group needing to further refine their propositions and demonstrate a clear development plan, whereas more mature firms tended to have a better understanding of market challenges and where innovation services could assist.

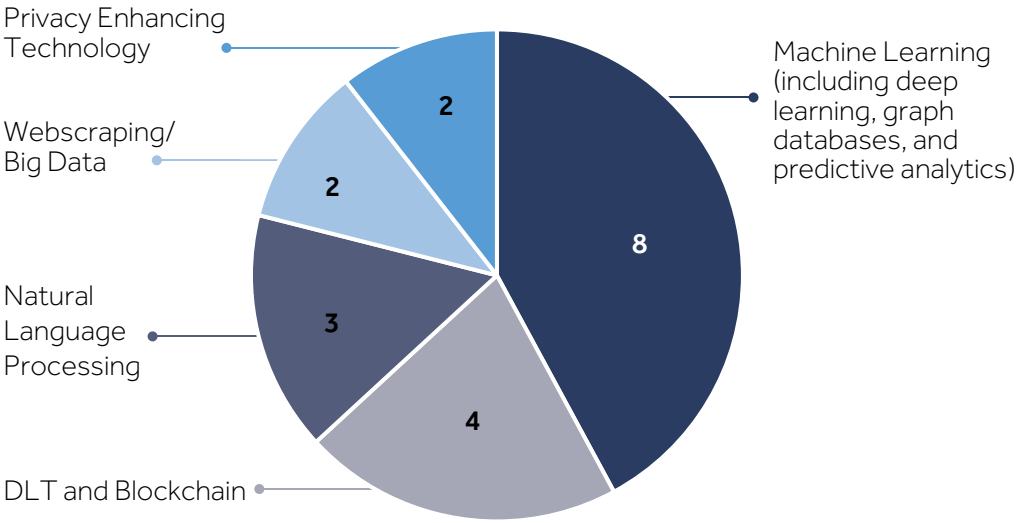
Figure 1 – applications and acceptance rate by firm size



2.6 Successful applications proposed several technologies and approaches that met the criteria for genuine innovation. The most common type of technology cited was machine learning – employed to make better use of the large volumes of disparate ESG data to drive actionable insights.

2.7 Blockchain and DLT were also popular, with immutability (trusted and verifiable data) and the decentralised nature of these technologies being important to the solutions. Natural Language Processing (NLP) was also core to three applications, with applicants deploying NLP to analyse large volumes of unstructured data that is common in the ESG sector.

Figure 2: frequency of core innovative technology underpinning applications



2.8 Solutions were varied and involved multiple innovative components but at a very high level three common themes emerged:

- Novel use of advanced analytics to collect and interpret existing data sources, improving the quality and volume of data ingestion (via NLP, webscraping, sentiment analysis, etc)

- Innovative use of technology and advanced analytics to analyse and produce insights in novel ways, such as through graphical databases, clustering and deep learning. A key focus was also on transparency of analysis, enabling verification and validation of outputs.
- Tapping into alternative data sources to enable a step-change in business models and approaches to existing products/services (satellite data, IoT data, etc), or verify existing ones.

2.9 A full list of successful participants, and a summary of the solution they tested during the sustainability pilot is available in **annex 1**.

3 Data

3.1 Financial data is highly sensitive and subject to data privacy laws that place conditions on sharing this data to protect the privacy of consumers.

3.2 To alleviate the data access challenge faced by many innovators, the FCA has explored the potential of synthetic data through our TechSprint and digital sandbox initiatives, as well as our own research.

3.3 The first digital sandbox pilot experimented with making synthetic data assets available to firms for the purposes of training machine learning models, and testing and validating products. In July 2020, the FCA hosted a three-week 'DataSprint' with over 100 industry participants prior to the launch of the pilot. The DataSprint produced reference data for millions of individuals and businesses, however the data was not tailored to the specific requirements of successful applicants. Participants of the first pilot expressed through feedback channels that the generic nature of the data was insufficient in some way for meeting their use case.

Sustainability pilot synthetic data generation

3.4 We adapted our approach for the sustainability pilot in favour of an iterative data generation process, tailored to each of the participants' requirements. Our partnership with Smart Data Foundry was central to this approach, who provided synthetic data expertise throughout the process of gathering requirements and generating synthetic data assets.

3.5 As part of the application process, we asked applicants to provide a high-level summary of additional datasets that would help them to develop their solutions. We received a range of requests in differing degrees of detail, from "geospatial data and satellite imagery" to "ESG labelled bond ratings & their methodology".

3.6 In November 2021, the FCA hosted a series of workshops with each team to build a detailed picture of their data requirements, for example whether they required real or synthetic data, the volume required, and what linkages needed to exist in the data.

3.7 Throughout January and February 2022, the FCA team collaborated with Smart Data Foundry to produce four synthetic data assets specifically tailored to the teams within the sustainability pilot. Throughout this period, we maintained communication channels with the teams to further iterate and improve the usability of the synthetic datasets.

3.8 Although some of the data assets were not available at the start, we believe that this iterative approach was beneficial as it produced more tailored results with improved usability in the long-term. Partnering with Smart Data Foundry also provided an opportunity for FCA data scientists to leverage external expertise and build our own knowledge of different synthetic data generation methodologies.

3.9 Despite the overarching benefits of this approach, we were not able to meet all of the requirements for each team. For example, one team required very specific data related to sustainability bonds that was neither openly accessible nor procurable within the timescales available. The challenges we experienced in this case highlight broader challenges in the ESG market, notably that the real data may not yet exist to test and develop specific solutions.

3.10 A summary of data assets made available to participants can be found in **annex 2**.

4 Summary of feedback

4.1 Over the course of the sustainability pilot, we asked participants to complete 3 feedback surveys, an exit interview, as well as attend monthly check-in sessions with the Working Group. These recurring touchpoints enabled a continuous feedback loop from which quick in-flight changes could be made. The exit surveys and interviews informed the evaluation of the sustainability pilot as well as future thinking around the digital sandbox. In addition, an independent evaluation of the sustainability pilot was undertaken by a third party.

Usage statistics from the platform

Users

4.2 The digital sandbox platform was available for team participants, expert mentors, and observers.

4.3 There were 57 unique users of the platform across the 12 participating teams with an average of 4.75 individuals per team, up slightly from the first pilot. Weekly active users (including observers) averaged 105.

4.4 The teams also utilised the platform's private messaging function to communicate with the Working Group, mentors and other teams in the sustainability pilot. In total, 1,540 chat messages were sent by the teams over the course of the pilot.

4.5 Over 500 viewers watched the live showcase sessions held in January and March, including representatives from financial services firms, national and international public bodies, academia, start-ups and technology firms. The final demonstration day videos can be viewed on each team's project page on the digital sandbox platform.

Data usage statistics

4.6 Users had two ways to access the data available on the digital sandbox platform: via API calls or by accessing datasets embedded in the platform and processing them using the integrated Jupyter notebook or off-platform. Aggregated usage statistics from the platform between January and end of March 2022 show significant engagement with the API and data features. In total there were:

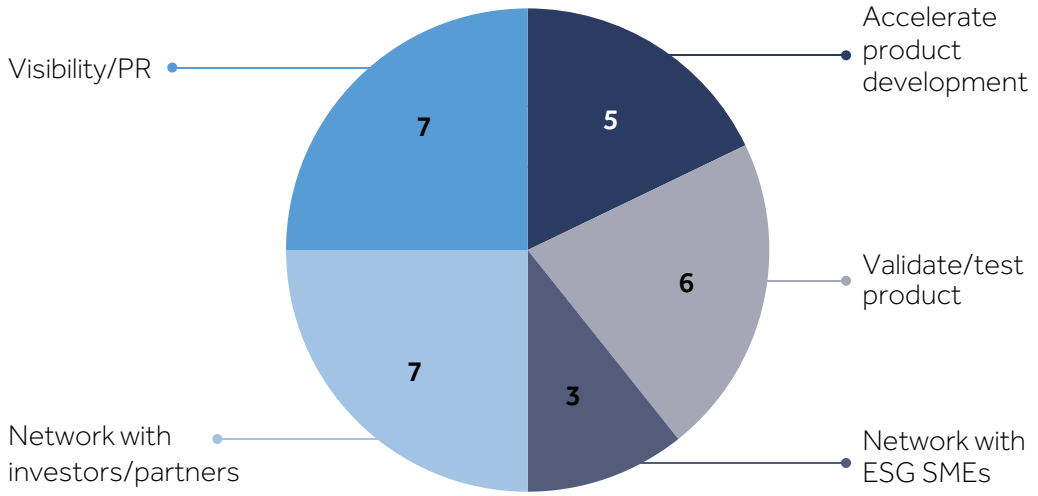
- 5,281 API calls
- 1,855 dataset queries

Initial survey summary

4.7 In terms of what participants hoped to gain from the sustainability pilot, the most stated answers related to networking with investors or partners, and visibility. 71% of participants hoped to broaden their network, develop partnerships and collaborate during the digital sandbox.

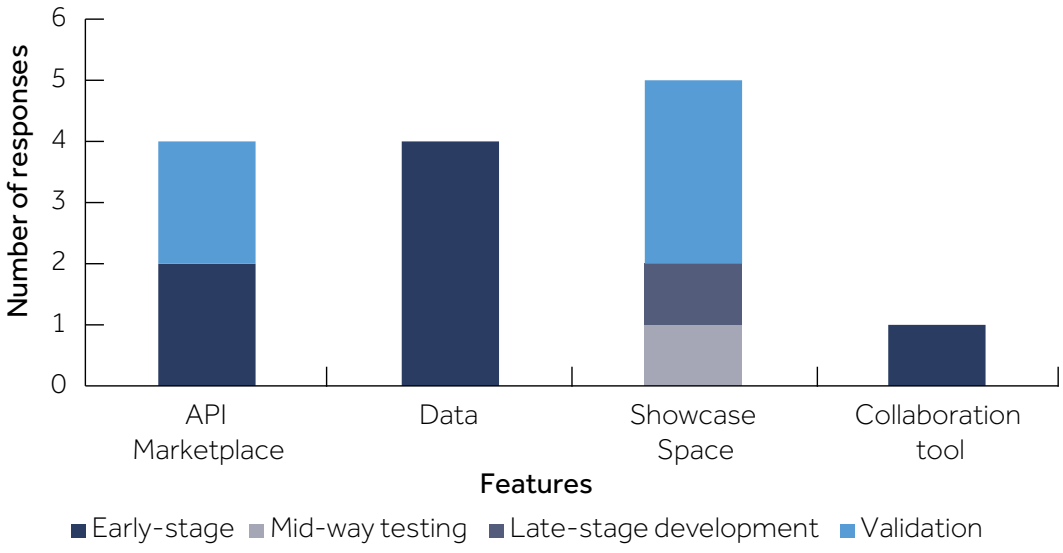
4.8 71% of respondents also hoped to accelerate their product development and/or validate and test their product during the sustainability pilot. Respondents who indicated product acceleration in this question were varied in terms of their stage of product development with some teams in early-stage testing and others at the point of validation.

Figure 1: what participants hope to gain from the sandbox



4.9 Six out of seven respondents whose solution was at 'early-stage' testing selected data-related features (API marketplace and data) as the most useful feature in the digital sandbox. Conversely, respondents at more advanced stages of product development tended to express more interest for the showcase space and the collaboration features. This pattern of responses is to be expected given that once a firm has tested and validated a solution, their attention will likely turn towards raising the visibility of their product and expanding a network of potential investors and partners.

Figure 2: most useful feature by product development stage



4.10 There was also a significant motivation to collaborate during the sustainability pilot, with 86% of respondents stating that they plan on working with other teams in their use case.

Exit survey and interview summary

Feedback on the digital sandbox initiative

4.11 All respondents to the exit survey indicated that the digital sandbox had a significant impact on accelerating innovation. This finding is to be expected given that we selected successful participants based on the premise that their offering was an example of novel innovation.

"ultimately it brings together people all working in this space and sharing that knowledge, ideas and sessions definitely encourages innovation".

4.12 Respondents widely regarded the theme of sustainability as pioneering for bridging the gap between FinTech and Sustainability. However, it was noted that the use cases were broad and at times overlapped, which was of particular detriment to participants who entered the digital sandbox with a high-level focus or without a clear objective in mind.

Feedback on collaboration

4.13 For the most part, respondents referred to the networking and collaboration features as valuable and 'one of the best aspects of the sandbox'.

4.14 Access to quality mentors was a key benefit for networking and increasing visibility and received excellent feedback from participants. Mentor and advisor engagement increased significantly in the later months of the sandbox:

- 50% of participants said they were actively engaging with mentors in December – this statistic increased to 100% in February.
- 30% of participants said they were actively engaging with investors/investment advisors in December – this increased to 73% in February.

4.15 Despite the initial survey indicating that teams were interested in working with other teams in the sustainability pilot, there were challenges in facilitating collaboration in this area, especially as the pilot progressed. By February, only 9% of teams stated that they were actively engaging with other members of the pilot.

Feedback on data

4.16 We received mixed feedback on the data assets, suggesting that the data provided more value to some participants than others. A proportion of participants stated that the APIs and datasets helped them to significantly advance product testing and validation, with a note that continued access to the data following the initiative would provide further benefit.

"Going through the different data sets opened our eyes to the different opportunities available".

4.17 However, a large proportion of participants suggested that the datasets required to validate their solutions were not available. Some overcame this by using publicly available data where possible, whereas others developed their own synthetic data to ensure they could still utilise the sandbox's network to receive user feedback and exposure.

4.18 Participants also stated that they would have preferred the data to be available at the start of the sandbox, rather than data being made available throughout the initiative.

Feedback on platform

4.19 A consistent theme revealed through the exit interviews and the platform usage statistics is that users preferred to extract the data and information on the platform and use it offline, rather than engaging with the features of the platform itself. This behaviour was particularly prevalent in both the data and messaging features.

4.20 All respondents mentioned that they preferred to download and integrate the data into their local machines rather than use the integrated Jupyter Notebook. This aligns with the behaviour of participants during the first pilot and suggests that the integrated environment provides limited benefit for participants.

4.21 Participants also tended to use the platform's messaging feature to initially reach out to mentors and other teams, but would subsequently take communications off-platform, with platforms such as Outlook or LinkedIn providing more convenience.

4.22 These findings suggest that both the data assets and the networking features have the most room for improvement and that these features can add significant value to participants in the sandbox.

4.23 Across the board, participants thought that the Working Group provided strong communication, facilitation and responsiveness. Those participants that had feedback during the sandbox felt that their requests were actioned effectively by the digital sandbox team, for example by connecting teams with mentors.

Independent evaluation

4.24 The independent evaluation assessed the sustainability pilot against five success criteria set at the start of the pilot:

- **Innovation:** role played in encouraging innovation in financial services
- **Market readiness:** role played in enabling the development and validation of proof of concepts and business model improvement.
- **Iterative development:** role played in developing a continuous improvement model to inform the design of a permanent digital sandbox environment.
- **Collaboration:** role played in fostering collaboration, facilitating diversity of thinking and creating an ecosystem of key stakeholders.
- **Thought leadership:** role played in developing the FCA's and other policy-making bodies' thought leadership on complex challenges and providing an evidence base for new policy.

4.25 Structured interviews were conducted with all participating teams, 2 advisory panel members, and 5 mentors. The table below provides a summary of the findings of the independent evaluation. A full circle in the 'status' column indicates a full score for that success criteria.

Table 1: high-level summary of independent evaluation findings

Success criteria		Status	Summary
Innovation	Role played in encouraging innovation in financial services		There was a significant impact on innovation acceleration, with the caveat that it is still too early to predict the full impact of the sustainability pilot on innovation.
Market readiness	Role played in enabling the development and validation of proof of concepts and business model improvement.		The digital sandbox accelerated product development for the vast majority of participants, however there were mixed views on the extent of business model improvement and product validation.
Iterative development	Role played in developing a continuous improvement model to inform the design of a permanent digital sandbox environment.		The digital sandbox provided a useful set of tools however access to relevant datasets is a key area for improvements and the platform itself could add much more value.
Collaboration	Role played in fostering collaboration, facilitating diversity of thinking and creating an ecosystem of key stakeholders.		The digital sandbox is an excellent way to network and provides the necessary collaboration across stakeholders to bridge idea, realistic use cases and capability to deliver.
Thought leadership	Role played in developing the FCA's and other policy-making bodies' thought leadership on complex challenges and providing an evidence base for new policy.		The pilot theme was pioneering for bridging the gap between FinTech and Sustainability, however there was limited evidence to suggest that the digital sandbox has/will impact policy development.

5 Lessons Learned

5.1 The following section details lessons learned and, where relevant, corresponding actions.

Use Cases

5.2 Although participants and mentors praised the focus on sustainability for this pilot, ESG data and disclosures is a broad theme that can at times be difficult to link with financial services.

5.3 A broad pilot theme invites significantly varied requirements for datasets and stakeholder networks. The ability to provide tailored datasets to all participants therefore suffered due to the sheer variety of requirements submitted across the different teams. Common themes existed across the requirements, including satellite imagery and ESG bond data, however we also received a large number of specific requests, for example data from ESG pension funds or granular metrics on corporate ecological impact disclosures.

5.4 Furthermore, ESG reporting and disclosure is a relatively new field in which open and rich datasets are not always accessible. We did experience cases where the data did not yet exist to meet a participant's requirements, thus requiring them to change the scope of their project. Although this might have hindered initial progress for some of the teams, discovering this issue at an early stage of the pilot prompted the teams to pivot their solutions to suit the data currently available in the industry.

5.5 When selecting the focus areas for the digital sandbox, due consideration should be made towards the data available in the field, to ensure that sufficient quality data exists to meet participants' testing requirements. Some participants also noted that they were unable to connect with key mentors, including from FinTechs and regulators. A more focused theme spanning fewer industries could help to narrow the scope of relevant mentors.

Synthetic data generation

5.6 As mentioned, the sheer breadth of requirements and the challenges in accessing real data, specifically data related to corporate ESG metrics or bonds, hindered our ability to produce datasets that met all participants' data requirements. It is also important to note that in addition to the synthetic data assets provided, teams were given access to other publicly available datasets. Where the synthetic data assets did not meet their requirements, some participants in the sustainability pilot utilised the publicly available datasets to develop their solutions, although these datasets were not bespoke or directly tailored to their requirements.

5.7 Ultimately, the approach used for the sustainability pilot did not deliver the finalised datasets required for all participants in the sandbox. Both the approaches used in the digital sandbox pilot and the sustainability pilot have their benefits and

limitations, and therefore future iterations of the sandbox will need to consider how to balance producing bespoke data sets in the timescales available, and whether it may sometimes be more appropriate to generate generalised synthetic data assets (not specifically tailored to participants' requirements) before the pilot starts. Considerations for the latter may include the number of firms in the sandbox, the breadth of requirements and the availability of real data.

5.8 Central to the data generation process during the sustainability pilot was the partnership with Smart Data Foundry. This partnership proved highly valuable in terms of leveraging Smart Data Foundry's synthetic data expertise to produce quality datasets for the participants. Exploring potential partnerships to improve the participant experience and the outcomes of the digital sandbox will form part of our thinking around future iterations of the initiative.

Inter-pilot collaboration

5.9 Collaborating with other teams in the sandbox is actively encouraged as a means of developing partnerships with firms at similar stages of development and also to progress proof-of-concepts by leveraging the experience and offerings of other teams. This form of collaboration may also lend itself to the generation of novel ideas and solutions, particularly where teams have complementary skillsets and products.

5.10 Participants may have benefited from a more structured approach to encouraging collaboration between different teams. If partnerships can be formalised earlier in the sandbox, this could help to overcome the challenges experienced with collaboration between teams at later stages of the pilot.

5.11 Future iterations of the digital sandbox could also promote collaboration by hosting more in-person activities where possible.

Alumni collaboration

5.12 A theme that emerged from the feedback was that continued access to the digital sandbox's platform and services would generate significant benefits for participants. In particular, participants stated that continued access to the data could further aid product development and validation once the sustainability pilot has concluded. Data is a core component of the digital sandbox initiative, and therefore we are currently exploring options to facilitate long-term access to the digital sandbox data for alumni, as well as opening up data access to a wider audience in GDPR-compliant way.

5.13 Sustained engagement with digital sandbox alumni could also further develop a thriving ecosystem of stakeholders for future digital sandbox participants. Digital sandbox alumni hosted several collaboration sessions during the sustainability pilot, and offered advice around preparing for the showcase days and their experiences since the pilot ended. Anecdotal feedback indicates that these sessions were well-received by current digital sandbox participants, suggesting that future sandbox participants would benefit from an active ecosystem of alumni to offer advice and mentorship.

Product development

5.14 During the first digital sandbox pilot, participants cited the synthetic data as the most valuable feature in enabling effective testing and product development. Interestingly, when asked the same question, the sustainability pilot participants referenced collaboration-related features, such as mentors, collaboration sessions and networking, as key accelerants.

5.15 This finding could indicate that where participants derived limited value from the datasets, they were able to leverage the other services of the digital sandbox to develop their solutions. Here, 'development' may not refer to product testing in the technological sense, but rather testing and developing the concept with mentors and advisers to ensure that the overall solution resonates with future investors and partners.

5.16 Future iterations of the digital sandbox should ensure a holistic offering of both quality data assets and collaboration features to complement product development from both technological and conceptual angles.

5.17 Ultimately, it is difficult to measure the full extent of product development over a limited period of 11 weeks. The Working Group should therefore follow-up with participants after 6 and 12 months to assess the extent of business acceleration, solution improvements and investment.

Platform

5.18 Both the usage statistics and feedback received from participants indicate that whilst some features of the platform provided value for users, the platform itself was not widely used over the course of the sustainability pilot.

5.19 There are undoubtedly benefits to maintaining user engagement with the platform, notably the monitoring of usage and ensuring sustained engagement with the pilot. Having detailed usage data could help to build a comprehensive picture of participant behaviour in the pilot; data that could subsequently be used to further develop and iterate the platform to serve user's needs.

5.20 However, the convenience of accessing the data outside the platform for example, where it can be used in conjunction with alternative data sources to develop prototypes in a local environment, will likely be preferred to any enhancements we could make to the platform to increase user engagement.

5.21 There are several short-term adjustments we could make to improve the user experience on the platform:

- providing alumni access to the platform to increase long-term engagement and further develop the ecosystem of stakeholders available to future participants.
- providing a more detailed 'profile' of each mentor including details of their role, areas of expertise, and company and contact details. For future iterations, mentors could provide these details on a mandatory basis during the application process. This information would make it easier for participants to identify useful parties to contact throughout the initiative.

5.22 In the long-term, the focus should be on developing the data and networking features on the platform, as these features derive the most benefit for participants. 'Success' in this area however should be evaluated against the quality of the datasets, the variety of mentors available and the quality of their interactions with participants, as opposed to targeted metrics on platform usage.

6 Next steps

6.1 The Digital Sandbox Pilot and Sustainability Pilot have demonstrated the ability for collaboration and access to data to stimulate beneficial innovation in the market. We are committed to establishing a permanent operating model for a digital testing environment that is based on the principles, and is improved by the experience and feedback, of the Digital Sandbox initiative. We are currently undertaking further research and industry engagement to ensure that this future service meets the needs of its end-users, and to understand the most relevant potential use cases that this initiative could support.

6.2 Beyond the application of a digital testing environment and collaboration platform to support market innovation and our existing regulatory services such as the Regulatory Sandbox and TechSprint programme, we are also exploring further potential use cases, such as facilitating regulatory-to-regulatory collaboration in, for example, the development of SupTech solutions.

6.3 We are also exploring methods to make the digital sandbox data assets openly accessible in a manner that complies with data protection laws. Datasets would be available as part of a public good designed to promote innovation and competition across financial services.

Annex 1
Successful application summary

Understanding ESG Data	
Connect Earth	A solution that aggregates and standardises environmental data on companies and products via an API. This can be embedded directly into the financial products that consumers and SMEs use every day. The focus is on giving consumers and SMEs insights into merchant- and product-level emissions data, to empower habit change towards more sustainable purchasing behaviour. Machine learning (graph networks) and k-nearest algorithms are used to standardise carbon emissions data and make company data comparable in a graph database.
Vested Impact	Focusses on 'impact' rather than traditional ESG by using a proprietary 'LIVES' scoring method. Ingests data from over 1,500 sources and uses advanced analytics and NLP to provide impact scores and comparable metrics for capital markets assets. Also provides impact ratings on over 100,000 private companies.
Butterfly Data	Customisable, consumer focused dashboards for ESG data. Initial MVP will focus on water footprint with a further roadmap of ESG metrics. Customisable dashboards will enable consumers to prioritise and weight their own preferences. Tech includes NLP, automated media scanning, ML of data sources. Designed to support the increasing cohort of retail investors using fintech to invest (eToro, Robinhood, etc).

Transparency and Disclosure	
Synectics Solutions	A proposal to create a standardised data model for the storage and transmission of ESG data, including compliance reporting, consultancy, risk assessment for potential fraud or misrepresentation, through an interoperable, open source data ecosystem. Enable organisations to understand self-attested ESG credentials and map out supply chains, associated partners, etc. to support ESG-related decision making through network analytics. A particular focus will be supporting sustainability-linked lending for financial services.
ESGi	A product, "ESGi" developed on DLT to enable direct sharing of ESG data between financial institutions who need to source information (for regulatory disclosure, risk, investment decision making etc.) and the companies who they need to assess. The initial use case supports SFDR data capture for Asset Managers to collect data from their Investees. Increases operational efficiency for firms with direct outreach to investees, clients and suppliers of ESG data.
QPQ	A proposal to develop a hybrid-data network, a system that interoperates with DLTs and database technologies. This enables proof of provenance, privacy & disclosure and big data storage while allowing analytics with privacy guarantees. The network enables ESG actors to exchange and transmit verifiable ESG measurements, reports and certificates with associated capabilities for custody, issuance, and exchange of ESG assets.
Coriolis Technologies	To develop a solution 'SusTrade', that will provide independent sustainability passports for companies and their supply chains. Banks can use these passports to verify their customers and continue to provide access to trade finance. Passport data will be populated from a) geolocation (BvD, OpenCorporates, GLEIF), webscraping, GIS APIs b) product and service analysis (HS/EBOPS Codes) c) media and social harvesting using GPT-3, as well as other data sources . Data will be correlated using deep learning and scores produced.

Validating ESG Data	
Greenomy	Aims to digitalise and build a standardised reporting framework for the EU Green Taxonomy, TCFD and upcoming UK Taxonomy. A systematic and digital portal will be created for companies (for report generation) and an audit portal (first line internal audit) with the following features: mandatory backing of responses in form of supporting documentation: audit trail to capture any changes and delay in responses; automatic cross-checking with supporting sources; automatic measurement of deviation between reported data, historical data, data of peers allowing the red flagging of potential outliers, mistakes or frauds; blockchain validation of certified ESG report and dissemination of reports; data capturing technologies such as IOT sensors to feed real-time data into reports, enhancing accuracy.
4 Earth Intelligence	Utilises an alternative data source – satellite-based Earth Observation (EO) to provide UK Asset Managers with new, regular or (near) real-time, independent information on the ongoing impact of their ESG bond portfolio. EO based monitoring and validation can be used to monitor both the project site itself as well as changes and impacts on the surrounding area (both natural and social impacts) automatically, through the combination of appropriate datasets and Machine Learning algorithms to extract information and quantify change.
Greenway Analytics	Aims to develop a solution that will connect real-time data from IoT devices to provide investors with a feed that can support validation of green bond KPIs. Their sandbox test will focus on Green Bonds issued by companies who want to reduce emissions in real estate by improving existing buildings. Initial IoT devices will use data from smart energy meters and thermal imaging devices to measure KPIs of green bond investments that focus on insulation
Arivu	A blockchain solution focusing on charge-of-state in ESG data. The solution aggregates, standardises and analyses disparate datasets into a user-friendly interface, and allows users to set automated triggers identifying key changes of state in those datasets (e.g. numeric value changes by 10%, field changes from 'Low' to 'High', etc.). Change-of-state data is then written to blockchain and indexed making immutable records accessible for audit and assurance as well as for statistical and business analysis of reported ESG datasets.
Telespazio UK	To develop a modular AI system called FinEO which uses satellite monitoring datasets to validate geographically and temporally specific asset disclosures from companies and corporations whose activities imply changes in land use that are detectable and can be monitored by satellite sensors. It will focus on the SASB Materiality Map which identifies sustainability issues, looking specifically at the validation of disclosures in the environment dimension and the ecological impact of Extractives and Mineral processing, providing accounting metrics on asset locations, areas of activity, areas of impact, areas of mitigation and areas of remediation.

Annex 2
Summary of data assets made available to participants

Dataset name	Type	Description
Consumer Transactions	Synthetic	A synthetic dataset containing 1.5m UK-based consumer transactions with accompanying account information for each consumer.
Asset Resolution Synthetic Dataset for Assets	Synthetic	Aggregation of asset-based activity indicators at the issuer level by sector and technology type.
Sustainability Linked Bonds	Synthetic	Contains green bonds, social bonds, sustainability bonds, and sustainability linked bonds.
Business Energy Usage Data	Synthetic	Synthetic representation of energy usage across a range of small, medium and large enterprises.
Cement Production Asset Level Data	Real	Cement production asset level data from the SFI Global Cement Database July 2021. This dataset provides details on cement plants this includes details on: the location of the plant, the owners, the stake of the owners, and the capacity of the plant.
European Red List 2017 December	Real	The European Red List is a review of the status of European species according to IUCN Regional Red List guidelines. It identifies those species that are threatened with extinction at the European level.
Iron and Stell Production Asset Level Data	Real	Iron and Steel Production Asset Level Data from the SFI Global Iron and Steel Database July 2021. This dataset provides details on iron and steel plants this includes details on: the location of the plant, the owners, the stake of the owners, and the capacity of the plant.
Planet Satellite Data	Synthetic	This synthetic dataset contains the metadata of the Planet Imagery Product.

Dataset name	Type	Description
Sustainable Development Goal Indicators	Real	UNSD indicator codes and the corresponding development goals. The 17 SDGs (sometimes called 'The GlobalGoals') is a collection of independent but interconnected goals carefully designed to give all of us on our planet a better future, with hundreds of targets and measurement indicators geared toward a date of 2030. The goals were created with businesses in mind, providing a path – what some people refer to as a "Pathway for Humanity" — for any business to harness their power by directing their efforts toward specific global objectives.
EU Red List Dataset	Real	The European Red List is a review of the status of European species according to IUCN regional Red Listing guidelines. It identifies those species that are threatened with extinction at the European level (Pan-Europe and the European Union) so that appropriate conservation action can be taken to improve their status.

Pub ref: 007847

12 Endeavour Square London E20 1JN
Telephone: +44 (0)20 7066 1000
Website: www.fca.org.uk