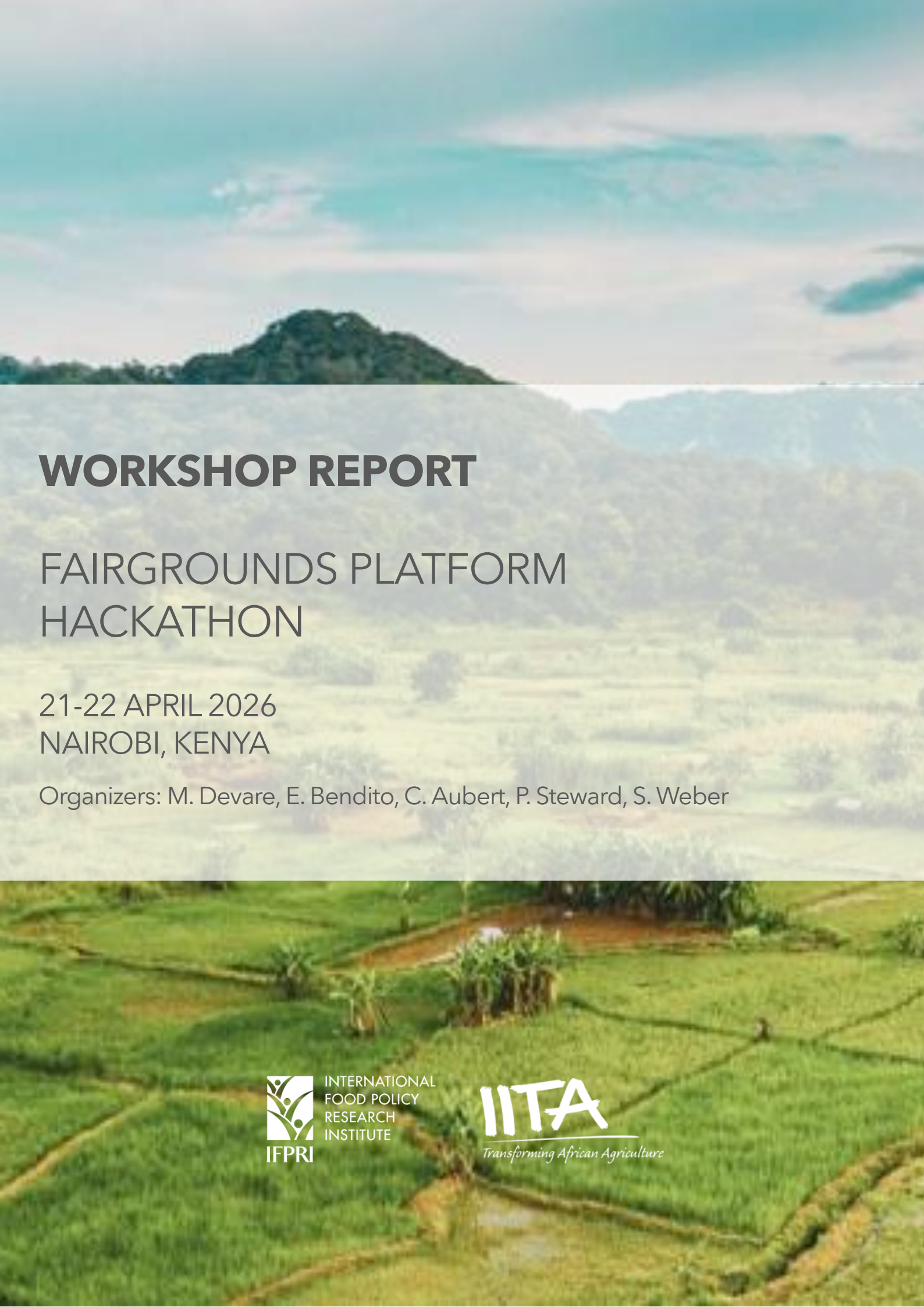


WORKSHOP REPORT

FAIRGROUNDS PLATFORM HACKATHON

21-22 APRIL 2026
NAIROBI, KENYA

Organizers: M. Devare, E. Bendito, C. Aubert, P. Steward, S. Weber

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI

IITA
Transforming African Agriculture

CONTENT

EXECUTIVE SUMMARY 3
PART ONE: INTRODUCTION 4
WORKSHOP OVERVIEW 4
PARTICIPANT PROFILE AND AFFILIATION 5
WORKSHOP METHODOLOGY 5
PART TWO: WORKSHOP CONTENT 6
ONLINE SESSIONS 6
IN-PERSON: SETTING THE STAGE 6
FAIRGROUNDS AND ITS GOVERNANCE 6
DATA & ANALYTICS : WHAT EXISTS 8
DATA MANAGEMENT - METADATA & STANDARDS 9
PART THREE: USE CASES 10
SPATIAL MODELING USE CASE – DOWNSCALING WAPOR DATASET 10
SPATIAL MODELING USE CASE – CROP TYPE MAPPING 10
CROP MODELING USE CASE- CROP MODELS AND YIELD PREDICTION 10
CROP MODELING USE CASE - EMPIRICAL MACHINE LEARNING 11
CLIMATE MODELING USE CASE - CLIMATE FORECAST FOR CROP & LIVESTOCK 11
PART FOUR: PARTICIPANT FEEDBACK 11
ONLINE SURVEY 11
FEEDBACK FROM USE CASES 14
PART FIVE: RESPONSES FROM FAIRGROUNDS TO PARTICIPANTS FEEDBACK 15
PART SIX: WORKSHOP EVALUATION 16
CONCLUSION 19

ANNEX 1. AGENDA 20
ANNEX 2. LIST OF PARTICIPANTS 23
ANNEX 3. WORKSHOP CONCEPT AND DEVELOPMENT 24
ANNEX 4. USE CASE PROJECT : DOWNSCALING WAPOR DATASET 26
ANNEX 5. USE CASE PROJECT: SPATIAL DATA CROP TYPE MAPPING 27
ANNEX 6. USE CASE PROJECT: CROP MODEL AND YIELD PREDICTION 28
ANNEX 7. USE CASE PROJECT: EMPIRICAL MACHINE LEARNING 28
ANNEX 8. USE CASE PROJECT: CLIMATE MODELING 29
ANNEX 9. IN-WORKSHOP FEEDBACK SURVEY RESULTS 30
ANNEX 10. USE CASE GROUP DISCUSSION FEEDBACK 43
ANNEX 11. PRE AND POST WORKSHOP ONLINE SURVEYS 47

EXECUTIVE SUMMARY

Fairgrounds is a FAIR, collaborative data platform initiated in 2025 by CGIAR. The platform is built on FAIR principles (findable, accessible, interoperable, and reusable) and aims to transform how agricultural datasets are accessed, shared, and applied across CGIAR and the broader global research community. Fairgrounds is funded by a Gates Foundation grant supporting smallholder farming innovations. It provides researchers with scalable computing resources and tools for data science and machine learning applications.

The Fairgrounds platform is currently under active development, with the goal of creating an integrated FAIR data environment for managing agricultural research data: combining data upload, annotation, tiered sharing permissions, and collaborative analysis. The current release represents a Minimum Viable Product (MVP), integrating RoboFlow's computer vision and data management capabilities and Databricks' computing and analysis infrastructure, along with other tools and models. At this stage of development, the MVP is not intended as a finished product, but as a working prototype designed to be tested and refined through real-world user engagement. The platform was tested via a hackathon involving 28 participants from six CGIAR Centers, with three private sector companies providing support. The hackathon focused on use cases including spatial, crop, and climate modeling, and was run in a combined online and in-person mode. The goal was for participants to develop prototypes and test the Fairgrounds MVP under real-world conditions for data management, usability, compute, and analytical capabilities.

Key agricultural data platforms such as GARDIAN (https://gardian.cgiar.org) and Carob (https://carob-data.org) were presented, along with plans for the Climate Data Hub. The CGLabs computing platform was also presented, along with a session on the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles and metadata. These sessions provided participants with broader context for the data ecosystem in which Fairgrounds will eventually operate.

The Fairgrounds platform capabilities were tested by five use case groups focused on: downscaling satellite data, crop type mapping, yield prediction, fertilizer response modeling, and translating climate forecasts to actionable advice. These concrete use cases highlighted benefits and challenges in data integration and analytical workflows.

Participant feedback was constructive and substantive. Users valued the platform's integrated "one-stop" research environment, scalable computing resources, and support for geospatial and machine learning workflows. At the same time, they identified areas requiring significant improvement before broader adoption: the onboarding experience and documentation need strengthening, the Databricks

integration requires simplification, R language support should be expanded, the interface needs refinement for intuitive use, and large data upload speeds need improvement. Participants also requested enhanced collaboration features and transparent governance and permissions management. Concerns about future platform costs were noted as a factor that could affect adoption.

The Fairgrounds technical team received this feedback as valuable development input, committing to addressing these priorities in upcoming sprints, including improved training, enhanced documentation, enhanced platform functionality (including RStudio integration), faster data handling, and clearer articulation of Fairgrounds' role within the broader toolset.

The hackathon successfully generated the user insight needed to guide the next stage of Fairgrounds' development. It concluded with broad interest in Fairgrounds as a resource for agricultural research and analytics, emphasizing that success depends on optimized user experience, comprehensive support, and sustainable operational models. These will be essential to realizing Fairgrounds' potential as development progresses.

PART ONE: INTRODUCTION

WORKSHOP OVERVIEW

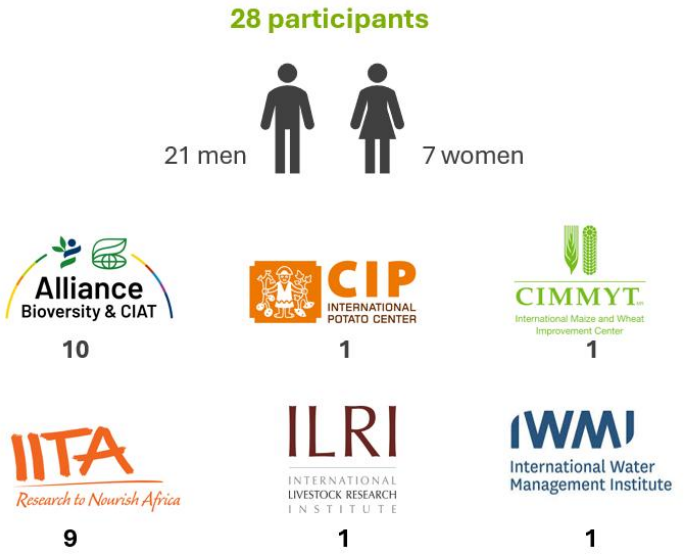
Fairgrounds is a platform to share, manage and use data collaboratively – from upload and annotation, to sharing based on tiered permissions, and analysis in collaboration with others. It is intended to offer data governance and guidance alongside powerful computing. To date, the platform brings together the RoboFlow computer vision and data management solution with Databricks for computing and analysis. Fairgrounds Proof-of-concept was developed with focus on digital phenotyping use cases; the April hackathon organized by IITA was conceived to test the current MinimumViable Product version of the platform for other agricultural use cases.

The workshop aimed to gather information on user needs regarding data and computing, and their experiences and feedback after testing and assessing Fairgrounds as a platform to offer these services. The objective was to assess suitability of the system to function as a common data hub and its analytical capabilities for different domains. The ability of the underlying Databricks Platform as a Service (PaaS) as a unified data and analytics platform to provide user friendliness and capacity to support analytical workflows was tested, with a report on fitness, ease-of-use and performance to advise on next steps for the Fairgrounds system with respect to data management and analytics.

PARTICIPANT PROFILE AND AFFILIATION

The workshop brought together stakeholders from the Digital Transformation Accelerator (DTA), Climate Action Science Program (CASP), Sustainable Farming Program (SFP), and Multi-Functional Landscapes (MFL) Program.

28 people participated in the workshop with 7 women and 21 men. 6 CGIAR Centers were represented: Alliance of Bioversity and CIAT, CIP, CIMMYT, IITA, ILRI and IWMI.



Three private sector companies also participated: Pariveda Solutions to provide tech support and training on Fairgrounds, Weber and Fritz Consulting as the primary consultant on the project, and SCiO Inc. who have been providing data ecosystem and computing support across projects and Science Programs at CGIAR.

WORKSHOP METHODOLOGY

The workshop included online and in-person hacking sessions. To test the Fairgrounds platform, participants worked with datasets to run analytical models on the platform and assessed its ability to support diverse data and analytics needs.

Participants were grouped by their expertise into crop, spatial and climate modeling use cases. These groups defined use case goals, identified key data and software requirements in advance, met prior to the workshop to develop solutions, and finalized their solutions and presented results in Nairobi. Pre-workshop sessions introduced the objectives and enabled the technical facilitators to provide training on the platform.

Following online trainings and trouble-shooting sessions with the Pariveda tech team, groups refined use cases, prepared datasets and tools, and developed their solutions.

During the workshop, time was split between group work and reporting on results, including experiences with data ingestion, software setup, analysis, modeling, and computing.

The overall objective was not to build new models but to test the platform end-to-end under tested prototypes, identifying strengths, limitations, and challenges. Feedback informed a recommendation report covering data management, usability, compute capacity, scalability, data sharing, cost, and sustainability.

PART TWO: WORKSHOP CONTENT

ONLINE SESSIONS

A series of online sessions was organized to introduce participants to the Fairgrounds platform and support the development of end-to-end data analytics or modelling prototypes.

The program began with an introductory session, which provided an overview of the Fairgrounds platform, including its core functionalities such as data discovery, ingestion, and transformation. Participants were guided through the use of catalog and workspace tools, as well as pre-configured datasets and computing environments. This session aimed to equip participants with the ability to independently navigate the platform and initiate their own analytical pipelines, with the expectation of delivering a 10-minute prototype demonstration during the in-person workshop.

Following the introduction, smaller use case team meetings were held to define project-specific objectives. Teams identified key research questions, expected outputs and required datasets, tools, and software. These requirements were communicated to the technical team to ensure integration and support within the platform.

Following the introductory session, teams conducted internal meetings to advance their work, assign responsibilities, and monitor progress. Two dedicated use case support sessions provided opportunities for troubleshooting, discussion of challenges, and technical assistance. During these sessions, participants also shared brief progress updates and received intermediate feedback on their use case work.

Overall, online engagement successfully guided participants from platform onboarding to use case prototype development, while providing continuous technical and collaborative support.

IN-PERSON: SETTING THE STAGE

To kick-start the in-person workshop, experts delivered presentations to introduce the overarching themes of the workshop, providing key concepts and platforms that would be covered during the rest of the workshop. The topics included Fairgrounds itself, existing tools, and the definition and importance of FAIR principles, metadata and data standards.

FAIRGROUNDS AND ITS GOVERNANCE

Fairgrounds is driven by the funding Gates Foundation and led by IFPRI to improve how agricultural research data are shared, discovered, and reused. It is embedded into the CGIAR Digital Transformation Accelerator initiative's agenda.
Fairgrounds serves as a federated gateway connecting Digital Assets across CGIAR

Centers and partner institutions without requiring data to be centralized. Digital Assets include datasets, code, models, pipelines, analytical tools (as e.g. RoboFlow and Databricks), documentation, metadata, and any other digital materials or resources. The platform is built around the FAIR principles: data should be Findable, Accessible, Interoperable, and Reusable. Fairgrounds aims to:

- improve interoperability across agricultural research systems,
- accelerate collaboration and innovation,
- support AI-ready and machine learning-compatible datasets,
- reduce duplication of research efforts, and
- strengthen open science practices.

Its infrastructure enables researchers to access and combine agronomic, climate, genomic, breeding, and socioeconomic Digital Assets through shared standards and cloud-based tools. The initiative supports CGIAR's Open and FAIR Data Assets Policy through a governance framework focused on trust, stewardship, and responsible sharing.

Key governance principles include:

- clear roles for data owners, stewards, managers, and users,
- institutional control over datasets within a federated system,
- common metadata and interoperability standards,
- protection of sensitive data and ethical reuse,
- scalable governance practices across CGIAR and partner organizations.

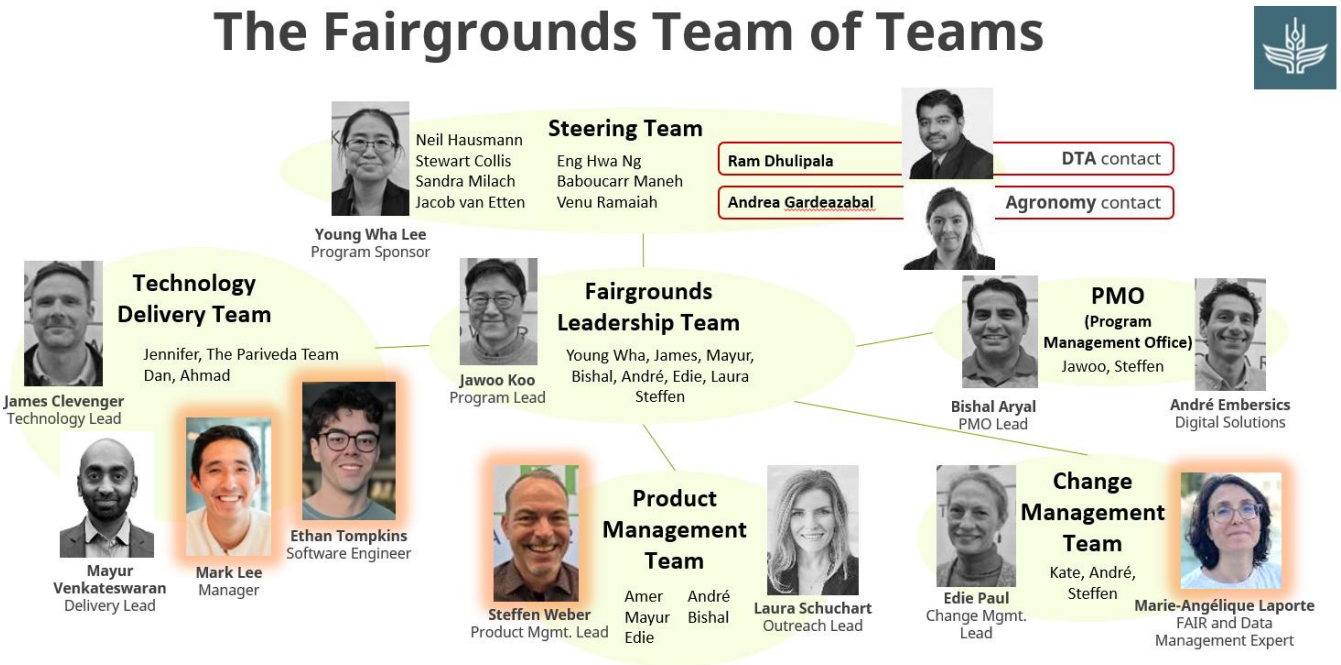


Figure 1: People involved in the Fairgrounds project. In orange : present at the workshop. Credit: Steffen Weber

Overall, Fairgrounds seeks to create a trusted, interoperable, and sustainable agricultural data ecosystem that strengthens collaborative research and evidence-based innovation.
Access the full presentation here, starting slide 2.

DATA & ANALYTICS : WHAT EXISTS

1. CLIMATE DATA HUB

The CGIAR Climate Data Hub is a shared, federated platform that connects and harmonizes CGIAR climate data assets through common standards and stewardship. It enables researchers and partners to use trusted climate evidence for research, policy, investment, and adaptation planning without duplicating data work.
The Hub supports both climate resilience and low-emission development through tools such as risk assessments and climate intelligence products.

2. GARDIAN

GARDIAN (https://gardian.cgiar.org) is a powerful digital platform designed to provide seamless access to thousands of data assets across CGIAR and beyond. GARDIAN brings together 26,000+ datasets, 470,000+ publications from 23 providers, and 570,000 standardized data records in one place, enabling users to discover, access, and reuse valuable agri-food research data often aligned with open science and FAIR data principles.

3. CAROB

Carob (https://carob-data.org/) produces large standardized data sets to facilitate agricultural research. It is an open-source, collaborative, and community based project supported by CGIAR. All data transformations are done with *R* scripts, making it easy to enhance the standardization process as needs arise, and to correct mistakes.

Access the full presentation here, starting slide 26.

4. CG LABS

CGLabs is a cloud-native, Kubernetes-based research and analytics platform developed by CGIAR and SCiO to support scalable scientific workflows. It provides researchers with on-demand computing resources, collaborative workspaces, and preconfigured environments such as Jupyter and RStudio for efficient data analysis. The platform integrates AI modeling, FAIR data infrastructure, and geospatial decision-support tools across agriculture, climate, and open science.

CGLabs offers a flexible and highly cost-efficient alternative to traditional cloud platforms, enabling advanced analytics with up to ~3× lower costs, predictable pricing, and optimized resource usage. The platform is evolving into an enterprise-grade solution supporting scalable, secure, and collaborative research across on-premise and cloud environments.

A hybrid storage model combines high-performance personal storage with scalable shared object storage to improve data management and collaboration. Shared workspaces, role-based access controls, support for external partners, and integrated interactive computing tools further strengthen collaboration while maintaining secure access.

The platform also includes a unified web portal for streamlined resource management, with future developments focused on cost-aware workload routing, dynamic cloud scaling, and improved monitoring and analytics. Overall, CGLabs provides a robust and flexible environment for reproducible and collaborative research.

Overall, CGLabs positions itself as a robust, flexible, and enterprise-ready platform for reproducible and collaborative research.

Access the full presentations here, starting slide 33 and slide 90.

DATA MANAGEMENT - METADATA & STANDARDS

The session introduced the concept of FAIR data and principles—Findable, Accessible, Interoperable, and Reusable—which support discovery, access, integration, and reuse by both humans and machines. FAIR data practices are particularly important in CGIAR, where data are distributed across Centers, projects, partners, and individual researchers, and are increasingly reused beyond their original context. Implementing FAIR principles helps reduce fragmentation, improve data reuse, and support automation, scalability, and innovation, including AI applications.

Fairgrounds supports FAIR principles. FAIRness is enhanced through metadata that improves findability, landing pages and APIs that support accessibility, the use of standards for interoperability, and provenance tracking, reproducible pipelines, and versioning that strengthen reusability. It was clarified that DOI assignment is not currently handled directly within Fairgrounds. Instead, Fairgrounds will provide landing pages containing descriptive metadata, citation information, and access links, with future DOI resolution expected to point to these pages.

Metadata and data standards are important to build FAIR data. Metadata standards are applied when registering datasets and assigning DOIs, while data standards address both the meaning and structure of data. Existing data standards include Terminag, CF Conventions, and ISO 19115, while structural standards include Carob, STAC, Zarr, COG, and Parquet.

Access the full presentation here, starting slide 43.

PART THREE: USE CASES

Participants tested Fairgrounds in groups, working on projects they had identified. Work was done online prior to the workshop, and 2 sessions allowed groups to finalize their projects during the in-person workshop. Below is a description of the 5 use cases.

SPATIAL MODELING USE CASE – DOWNSCALING WAPOR DATASET

Participants: Ambica Paliwal, Oscar Bautista, Zolo Kiala

The project aims to enhance the spatial resolution and usability of the WaPOR dataset by developing methods for downscaling geospatial data. The objective is to generate more detailed, locally relevant information to support agricultural monitoring, water resource management, and decision-making processes.

By refining coarse-resolution satellite-derived data into higher-resolution outputs, the project seeks to improve the accuracy of analyses at field and sub-national levels. This work is particularly relevant for applications requiring precise spatial insights, such as irrigation planning, crop assessment, and resource optimization.

Overall, the project contributes to making WaPOR data more actionable and accessible for researchers, policymakers, and practitioners working on land and water management (see ANNEX 4, and full presentation here, starting slide 65).

SPATIAL MODELING USE CASE – CROP TYPE MAPPING

Participants: Gilbert Rozario, Francis Muthonis, Eduardo Bendito, Bester Mudereri

This project aims to improve crop type mapping in Ghana's smallholder mixed farming systems, where reliable in-season crop statistics are limited.
Current field data collection is labor-intensive, while cloud cover, trees, and weeds make remote sensing classification difficult. The project will develop a machine learning model with high classification accuracy (target F-score >0.85) for identifying crop types.
The resulting crop maps supports site-specific extension services and help traders identify where crops are grown for demand aggregation.
Ultimately, the solution should improve the timeliness and accessibility of crop statistics for food security planning and policymaking (see ANNEX 5 and full presentation here, starting slide 70).

CROP MODELING USE CASE- CROP MODELS AND YIELD PREDICTION

Participants: Namita Joshi, Andres Aguilar, Siyabusa Mkuhlani

This project addresses the difficulty of predicting crop yields, which vary widely across locations, seasons, soils, and climate conditions. It uses the DSSAT model combined with climate, soil, and observed yield data to improve spatial yield prediction.

The solution generates consistent, location-specific yield estimates and helps explain yield differences across environments.
It enables users to predict yields, compare management scenarios, and identify high- or low-potential production areas. It supports better targeting of interventions and provides a scalable tool for use across regions and seasons (see ANNEX 6 and full presentation here, starting slide 80).

CROP MODELING USE CASE - EMPIRICAL MACHINE LEARNING

Participants: Caroline Owuor, Amsalu Tilaye, Sehou Romaric Kouiho, Degefie Tibebe

This project estimates common bean yield in Rwanda under different NPK fertilizer rates. It addresses low yields due to unbalanced fertilizer use and strong spatial variability in growing conditions. It uses a machine learning model to predict yield responses from trial and environmental data.
The model helps identify optimal fertilizer rates and quantify nutrient–yield relationships. It supports evidence-based, site-specific fertilizer recommendations for farmers and practitioners (see ANNEX 7 and full presentation here, starting slide 83).

CLIMATE MODELING USE CASE - CLIMATE FORECAST FOR CROP & LIVESTOCK

Participants: Lizeth Llanos, Diego Agudelo, Jemal Ahmed, Peter Steward

The project addresses the mismatch between seasonal climate forecasts and the shorter timescales at which farmers and pastoralists make decisions. This limits their ability to manage rainfall variability, drought, flooding, and livestock heat stress. The solution converts seasonal forecasts into actionable monthly climate risk information, including rainfall onset forecasts, drought/wetness indicators (SPEI), precipitation anomalies, and livestock heat stress maps. Integrated into digital advisory platforms, these tools support locally relevant crop–livestock recommendations. This enables farmers and advisors to make proactive decisions on planting, crop selection, irrigation, and livestock management, improving resilience, productivity, and resource-use efficiency (see ANNEX 8 and full presentation here, starting slide 86).

PART FOUR: PARTICIPANT FEEDBACK

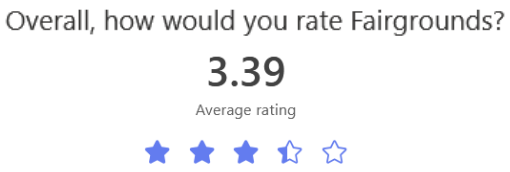
Participants shared their feedback on their use of Fairgrounds via an online survey and a use case group discussion guided by questions.

ONLINE SURVEY

18 participants provided their feedback using the online survey developed on Microsoft Forms. Below is the summary of the key points, full results are available in ANNEX 9.

Overall Impression

Respondents generally viewed Fairgrounds as a promising platform with useful data storage and integration capabilities, but many felt that its purpose, workflows, and overall user experience are still not fully mature or clearly communicated.

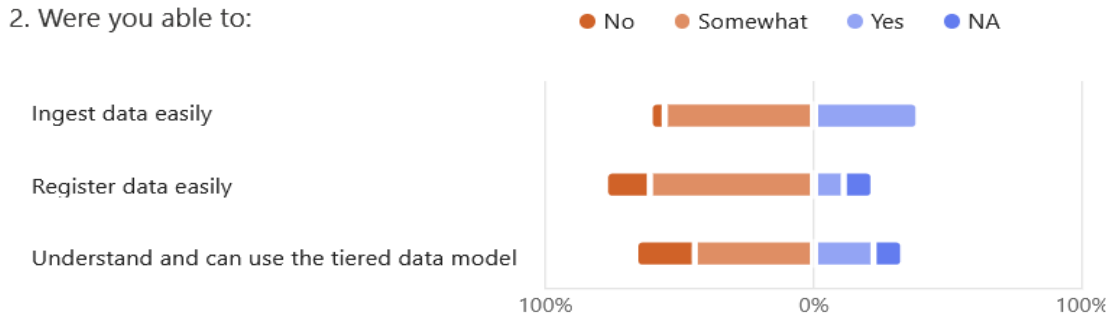


Data Discovery & Usability

Users appreciated the availability of search and filtering tools, but many reported difficulties locating relevant datasets due to unclear organization, incomplete metadata, and inconsistent search results.

Data Ingestion & Metadata

While most respondents were able to ingest and register data with some success, many found the process insufficiently intuitive and expressed a need for clearer metadata standards, simpler workflows, and better guidance on the tiered data model.



Governance & Permissions

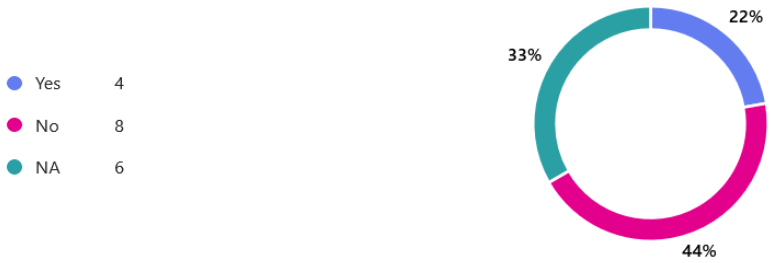
Survey participants highlighted governance, access control, and data stewardship as important concerns, with several users reporting confusion or frustration around permissions management and authentication processes.

17. What data governance considerations would you like to see improved in Fairgrounds?

Documentation & Training

A strong recurring theme was the need for improved documentation, practical tutorials, and beginner-friendly onboarding materials, particularly for users working in R rather than Python.

25. Was the guidance/documentation available sufficient to enable you to configure analytics needs easily?



Analytics Environment & Tooling

Users considered flexible software environments and package installation capabilities to be critical, but many encountered limitations related to library management, environment configuration, and support for advanced analytical workflows.

Databricks vs Fairgrounds

Many respondents interacted primarily with Databricks and felt that the relationship between Fairgrounds and Databricks remains unclear, leading to uncertainty about the distinct role and added value of the Fairgrounds platform.

Geospatial & Large Data Needs

A significant portion of respondents work with geospatial and large-scale datasets and emphasized the need for stronger spatial analysis tools, improved handling of large files, and integration with platforms such as Google Earth Engine.

Technical Skills Required

The survey suggests that effective use of Fairgrounds currently requires moderate to advanced technical expertise, including familiarity with Python, cloud computing environments, and data infrastructure concepts.

Key Recommendations

Overall, respondents recommended improving onboarding and documentation, simplifying workflows, strengthening governance and permissions management, enhancing R and geospatial support, and clarifying the long-term vision and functionality of the platform.

FEEDBACK FROM USE CASES

Participants also provided feedback during in-person discussions guided by a common set of questions across all use case groups. Below is a summary of the feedback from all use case groups. The full results are available in ANNEX 10ANNEX 9

Fairgrounds was widely recognized as a platform with strong potential because it combines data management, analytics, and computing in one platform. However, users consistently highlighted usability and workflow efficiency as the main areas needing improvement.

- Simplifying the user experience. Participants found the interface difficult to navigate, especially for new users, and requested easier file uploads, folder management, data registration, and dataset organization. Users also emphasized the need for clearer documentation and better visibility of available tools and features.
- Faster uploads and support for larger files without needing developer intervention were repeatedly mentioned as priorities.
- Computing performance and flexibility. Several users experienced failed simulations or slow workflows due to limited computing resources and complicated environment setup. They requested more computing power, simpler cluster and package administration, and better serverless performance for machine learning workloads.
- Support for R emerged as one of the most critical gaps. Many agricultural researchers primarily work in R, yet Fairgrounds currently lacks RStudio integration and easy R setup. Users strongly recommended adding RStudio and enabling automatic code translation between Python and R to improve accessibility and interoperability.

- Collaboration and reproducibility were also important concerns. While basic collaboration was considered straightforward, users requested GitHub-like script versioning, better tracking of modified datasets, and easier sharing of outputs with external, non-platform users.
- Participants additionally highlighted the need for stronger integrations and tooling, including APIs to external repositories, smoother DataBricks connectivity, geodata support, local IDE integration, and preinstalled scientific tools such as DSSAT.
- Compared with existing approaches such as CG Labs, Fairgrounds was valued for offering an integrated "one-stop-shop" environment for data and analytics. However, users felt competing platforms currently provide a more intuitive and streamlined experience.

Overall, the feedback shows that users see significant value in Fairgrounds, but broader adoption will depend on improvements in usability, documentation, computing scalability, R support, and workflow integration.

PART FIVE: RESPONSES FROM FAIRGROUNDS TO PARTICIPANTS FEEDBACK

During the workshop, the Fairgrounds technical team addressed participant feedback and questions regarding the platform. The Fairgrounds team responded to concerns related to training, platform functionality, governance, and future development plans.

Participants expressed strong interest in additional modular training sessions covering topics such as data registration and ingestion, workspace navigation, the tiered system, RStudio connectivity, Databricks connectors, and the Fairgrounds R support. In response, the team committed to expanding training resources, improving FAQs, and providing clearer guidance on platform use, data management, governance, coding systems, support requests, and public data sharing.

Several improvements were also discussed based on user feedback. Planned developments include enabling easier public data sharing, exploring dashboard functionality, improving third-party connectors, increasing upload capacity and speed for large datasets, simplifying file management, and making code and scripts more accessible to non-programmers.

Participants additionally raised questions about data governance, integration with external tools and repositories, discoverability of CGIAR data, and the relationship between Fairgrounds and Databricks. The team acknowledged the need for clearer platform boundaries and confirmed that several suggestions and integration possibilities will continue to be explored in collaboration with stakeholders and users.

Costing model and perspective were presented to participants, which sparked an engaged discussion. Participants were informed that, following the development phase, the platform's operational costs would need to be covered by users. This raised concerns across the room and affected participants' willingness to use the platform if it were not offered free of charge.

Access the full presentation here, starting slide 108

PART SIX: WORKSHOP EVALUATION

A survey was sent to participants before the first online session and at the end of the workshop. The pre-workshop survey was completed by 21 participants and the post-workshop survey was completed by 17. Link to full results are in ANNEX 11. The survey assessed participants' understanding of the platform, perceived relevance to their work, technical confidence, support experience, and recommendations for improvement.

Feedback from participants was generally positive. Most respondents saw clear potential for Fairgrounds in their research and workflows, especially for geospatial analysis, large-scale computing, machine learning, and collaborative analytics.

The workshop allows participants to gain understanding and confidence using the Fairgrounds platform. Following the Hackathon, all participants consider Fairgrounds relevant for the work.

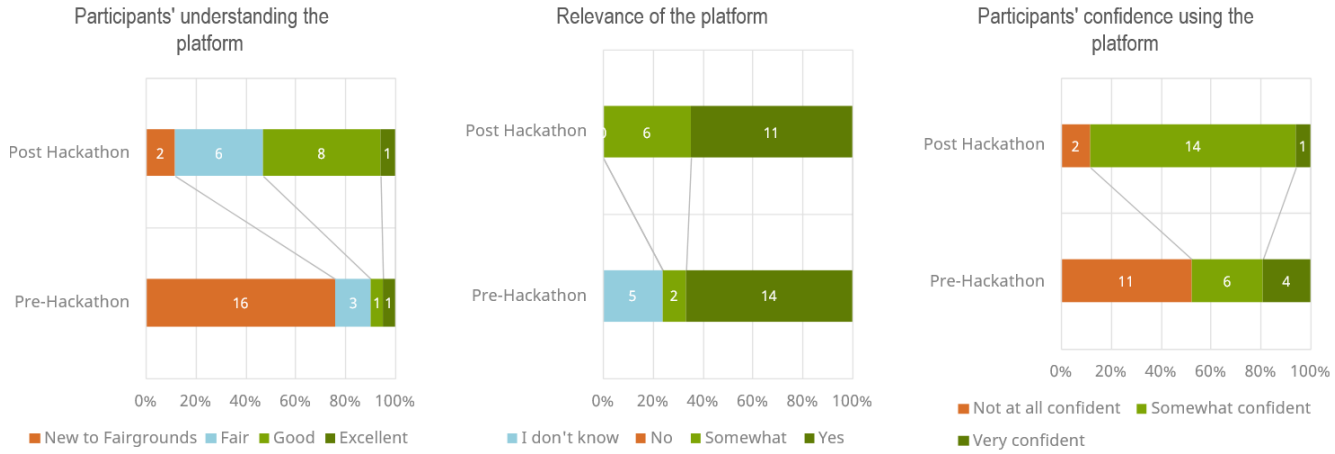


Figure 2: Survey results diagrams from Steffen Weber

When considering the potential for Fairgrounds to provide value beyond the value that participants currently receive from other systems, 15 out of 17 participants (88%) answered affirmatively.

15. Assuming you are provided sufficient support, do you see the potential for Fairgrounds to provide value to your work beyond the value you currently receive from other systems?

17 Responses

ID ↓	Name	Responses
17	anonymous	As long as Databricks is cheap, it's a great tool for work sharing.
16	anonymous	Databricks is tremendously useful and a powerful tool with huge potential for accelerating our processes. Although we haven't fully explored the connection with external datasets and the Data Marketplace yet, the possibilities are very interesting. Additionally, the opportunity to collaborate by viewing others' progress in shared notebooks and job processes is extremely valuable.
15	anonymous	Yes
14	anonymous	Not really. It is important Fairground put forward a clear set of capabilities it proposed rather than trying to solve problem from different use cases.
13	anonymous	Yes
12	anonymous	Yes
11	anonymous	Absolutely yes
10	anonymous	Yes, provided the features that we got access to are not removed.
9	anonymous	Yes, databricks can provided support in procesing and modeling
8	anonymous	Yes
7	anonymous	Yes
6	anonymous	yes
5	anonymous	Yes
4	anonymous	Compared to other systems, not really since most of the features are already available.
3	anonymous	Yes, but it required a lot of problem solving first- esp the R part
2	anonymous	Yes
1	anonymous	yes

14 participants over 17 felt they received sufficient support and 13 reported acquiring new skills. 14 participants said they were “ready to use Fairgrounds with support”.

14. How technically confident do you feel about using the Fairgrounds platform today?

Participants highlighted several platform capabilities as particularly useful:

- Access to scalable computing resources
- Geospatial data integration
- MLflow/model tracking
- Serverless compute
- Collaboration capabilities
- Analytics and workflow execution
- Preconfigured environments and packages

Although the overall experience was positive, participants consistently identified several challenges:

- **Databricks Complexity and Navigation.** Many participants found the integration between Fairgrounds and Databricks difficult to understand initially.
- **Need for Better Documentation and Tutorials.** More onboarding materials, video tutorials and step-by-step workflows
- **R Language Support.** Multiple participants noted challenges using R workflows within the platform, especially around environment setup and execution.
- **Long-Term Support and Infrastructure Concerns** Participants raised concerns regarding: Availability of continued technical support, Compute costs and resource allocation, Dependence on Databricks infrastructure, Integration with other systems and APIs.

As shown in the diagrams below, these challenges may have affected participants' willingness to use and recommend Fairgrounds. These findings should be interpreted with the understanding that the current version of Fairgrounds is a working prototype, currently under active development and designed for iterative testing and refinement.

18. How likely are you to utilize Fairgrounds for your future work and/or research?

Supporters	3
Passifs	7
Détracteurs	7

19. How likely are you to recommend Fairgrounds to a colleague or researcher?

Supporters	2
Passifs	8
Détracteurs	7

CONCLUSION

1. **Strong interest and perceived value:** Participants generally believe Fairgrounds shows potential to support research and analytics workflows, especially in geospatial and machine learning applications.
2. **Support was appreciated:** Most participants felt adequately supported during the workshop and valued direct interaction with facilitators and technical teams.
3. **The learning curve remains significant:** The largest barrier to adoption is the complexity of the underlying Databricks ecosystem and platform navigation.
4. **Training and onboarding are critical:** More structured tutorials, examples, and beginner-friendly documentation would significantly improve user experience.
5. **Participants are willing to continue using the platform:** Most respondents indicated moderate to high likelihood of using and recommending Fairgrounds in the future, provided sufficient support remains available.
6. **Optimal timeframe for such exercise:** Users felt 2 days was too short, and 3 days would be better for such an exercise.

The workshop successfully introduced participants to Fairgrounds in its Minimum Viable Product (MVP) stage, and demonstrated the platform's potential for research and analytics. While users appreciated the platform's computing power and geospatial capabilities, broader adoption will depend on Fairgrounds' continued development as it moves through further user feedback testing cycles and iterative improvements, addressing usability, onboarding, documentation, long-term support and pricing. Overall, the feedback indicated interest in continued engagement with Fairgrounds and optimism about its future applications.

ANNEXES

ANNEX 1. AGENDA

ONLINE SESSIONS

Date	Time	Description	Responsible	Objectives
March 30 (1530 EAT)	2 h	**Introduction to workshop** • Overview of Fairgrounds (15 min) • Use cases (climate, crop, spatial, data/compute) • Fairgrounds/Databricks use • 4. Example-driven training on FG use	Steffen Weber Eduardo Bendito, Peter Steward, Medha Devare Pariveda	Become familiar with FG platform (starting with recording); enable independent navigation of system
April 1-3	0.5 h /use case	**Use case team meetings to clarify requirements:** Agree on joint questions; expected hacking outputs for each use case Identify required datasets, tools, software (w/in 1 week of meeting) *Install required data & software needs (w/in 1 week of meeting)*	Eduardo Bendito, Peter Steward *Pariveda to help if/as needed*	Agree on critical questions & outputs for each use case; data and software requirements per use case so tech team can integrate as/if needed (e.g., R Studio, gdal)
April 7 & 14	1.5 h	**Use case sessions for teams to:** Discuss progress/issues/challenges Troubleshoot, get support	Pariveda, use cases	Troubleshoot and assist
April 1-17	2 h +	Use case teams set recurring meetings to advance on data & analytics	Use cases	Agree internally, assign responsibilities, plan, check progress regularly

DAY 1: APRIL 21, 2026

Time	h	Description	Who - How	Objectives
0830 - 0930	1 h	**Intro to workshop** 1. Welcome, intros across group 2. Working with Fairgrounds, how/why of user input & engagement; FG governance 3. Initial feedback from survey	Medha Devare Steffen Weber	Develop shared understanding of Fairgrounds & its governance model; air user concerns/ questions
0930 - 1030	1 h	**Data & Analytics: What exists** 1. Current solutions and why they exist (Climate data hub; GARDIAN - Carob) 2. CG Labs	Medha Devare, Peter Steward Pythagoras Karampiperis	Understand current state
TEA/COFFEE BREAK				
1045 – 1115	0.5 h	**Data Management: Metadata & standards** 1. Metadata & FAIR: what, why, how 2. Existing standards 3. Lessons learned re: Carob approach to standards	Marie-Angélique Laporte	Understand basics of data management, with focus on standards and lessons learned
1115 - 1300	1.75 h	**Use case final touch-base:** Time to discuss work, finalize presentations	Use cases	Finalize presentations – for each use case
LUNCH BREAK				
1345 - 1415	0.5 h	**Survey on users' experience using Fairgrounds**	Medha Devare	
1415 – 16 :30	0.75 h	**Groupwork - Data Management and Analytical Workflows**	Moderated UC groups with rapporteur	
TEA/COFFEE BREAK				
1645 – 1715	0.75 h	**Report back - data**		
1715 - 1730	0.25 h	**Recap; looking forward to tomorrow**		Review the day; overview of next day
19h30		**Social dinner**		

DAY 2: APRIL 22, 2026

Time	h	Description	Responsible	Objectives
0830 - 0900	0.5 h	Overview of day's program	Medha Devare	Set the stage
0900 - 1000	1 h	**Responses to group work feedback**	Mark Lee, Ethan Tompkins	Tech team views on user feedback
1000 - 1200	2 h	**Use case final touch-base**	Use cases	Finalize presentations – for each use case
TEA/COFFEE BREAK				
1200 - 1300	1 h	**Use Case outputs & experiences:** FG workflows, incl approaches to: 1. Data ETL – prep for analysis 2. Modelling, analytical workflows 3. Registering & sharing data & model outputs (incl. licenses etc.)	3 use cases (15 + 5 min Q&A) (slides and/or live demo)	Provide insight on work done, FG end-to-end workflows (data ingestion & management, analytics & modelling, and final outputs registered in the platform)
LUNCH BREAK				
1415 - 1500	0.75 h	**Use Case outputs & experiences:** FG workflows, incl approaches to: 1. Data ETL – prep for analysis 2. Modelling, analytical workflows 3. Registering & sharing data & model outputs (incl. licenses etc.)	2 use cases (15 + 5 min Q&A) (slides and/or live demo)	Provide insight on work done, FG end-to-end workflows (data ingestion & management, analytics & modelling, and final outputs registered in the platform)
TEA/COFFEE BREAK				
1515 - 1545	0.5 h	**SCiO reflections and next steps for CG Labs**	Pythagoras Karampiperis	Reflect on CG Labs sustainability/support model, costs, next steps…
1545 - 1630	0.75 h	**FG tech / product team reflections**	Steffen Weber	Reflect on FG sustainability/support model, costs
1630 - 1715	0.75 h	**Attendee outcomes survey**	Steffen Weber	
1715 - 1730	1 h	**Next steps**		

ANNEX 2. LIST OF PARTICIPANTS

Name	Affiliation	Role
Namita Joshi	Alliance Bioversity-CIAT	Testing the platform
Degefie Tibebe	Alliance Bioversity-CIAT	Testing the platform
Jemal Ahmed	Alliance Bioversity-CIAT	Testing the platform
Lizeth Llanos	Alliance Bioversity-CIAT	Testing the platform
Andres Aguilar	Alliance Bioversity-CIAT	Testing the platform
Peter Steward	Alliance Bioversity-CIAT	Organizer + Testing the platform
Aniruddha Ghosh	Alliance Bioversity-CIAT	Testing the platform
Diego Agudelo	Alliance Bioversity-CIAT	Testing the platform
Benson Kenduiywo	Alliance Bioversity-CIAT	Testing the platform
Marie-Angelique Laporte	Alliance Bioversity-CIAT	Representing Fairgrounds
Gilbert Rozario	CIMMYT	Testing the platform
Bester Mudereri	CIP	Testing the platform
Amsalu Tilaye	IITA	Testing the platform
Siyabusa Mkuhlani	IITA	Testing the platform
Caroline Owuor	IITA	Testing the platform
Francis Muthoni	IITA	Testing the platform
Oscar Bautisa	IITA	Testing the platform
Sehou Romaric Kouhio	IITA	Testing the platform
Celine Aubert	IITA	Organizer
Medha Devare	IITA	Organizer
Eduardo Bendito Garcia	IITA	Organizer + Testing the platform
Ambica Paliwal	ILRI	Testing the platform
Zolo Kiala	IWMI	Testing the platform
Mark Lee	Pariveda Solutions	Representing Fairgrounds
Ethan Tompkins	Pariveda Solutions	Representing Fairgrounds
Sotiris Konstantinidis	SCiO	Representing CG Labs
Pythagoras Karampiperis	SCiO	Representing CG Labs
Steffen Weber	Weber and Fritz	Representing Fairgrounds

ANNEX 3. WORKSHOP CONCEPT AND DEVELOPMENT

Objective

The workshop aims to gather information on user needs regarding data and computing, and their experiences and feedback after testing and assessing Fairgrounds as a platform to offer these services. The objective is to assess suitability of the system to function as a common data hub and its analytical capabilities for different domains. The ability of the underlying Databricks Platform as a Service (PaaS) as a unified data and analytics platform to provide user friendliness and capacity to support analytical workflows will be tested, and a report on fitness, ease-of-use and performance will be the output to advise on next steps or the Fairgrounds system with respect to data management and analytics.

Proposed Design

The workshop will involve stakeholders from the Digital Transformation Accelerator (DTA), and the Climate Action Science Program (CASP), Sustainable Farming Program (SFP), and the Multi-Functional Landscapes (MFL) Program. Participants will be tasked with managing and using data (including climate, hydrology, and other agronomy domains). They will use these datasets to run analytical models and workflows hosted on Fairgrounds and evaluate the platform's ability to support diverse data and analytics needs.

We propose the design outlined below, with responsibilities outlined in Figure 1:

- Participants will be grouped by use cases around tools/topics of relevance to their typical workflows (e.g. crop models, agronomy ML pipelines and/or climate analysis).
- Through these groups, participants will identify key requirements (datasets and software) needed to implement workflows so that data and analytical models can be ingested/installed in the hackathon/group workspaces well before the workshop.
- **Please note:** *Use case groups will need to meet a few times before the in-person meeting and develop agreed analytics solutions to test the platform's data, analytics and computing capabilities, and be prepared to report on their work in Nairobi*.
- Pre-workshop sessions will be organized to present the objectives of the workshop/hackathon, and prepare attendees on the platform.
- During these sessions technical facilitators of the Fairgrounds platform will guide and train participants, providing them with an in-depth introduction to the platform, set-up and features, with illustrative examples and time for in-depth Q&A and discussion.
- Following online trainings, users will spend time agreeing on specific use cases (1-2 per group), familiarizing themselves with the system, identifying needed data and software, performing ingestion/installation tasks and work in teams on their use case solution.

- During the workshop, participants will have some time to work in their groups, but much of the time will be used to report on their results and their experiences with Fairgrounds processes (data ingestion, software and library installation, data exploration and EDA, modelling and reporting tools) and computing.

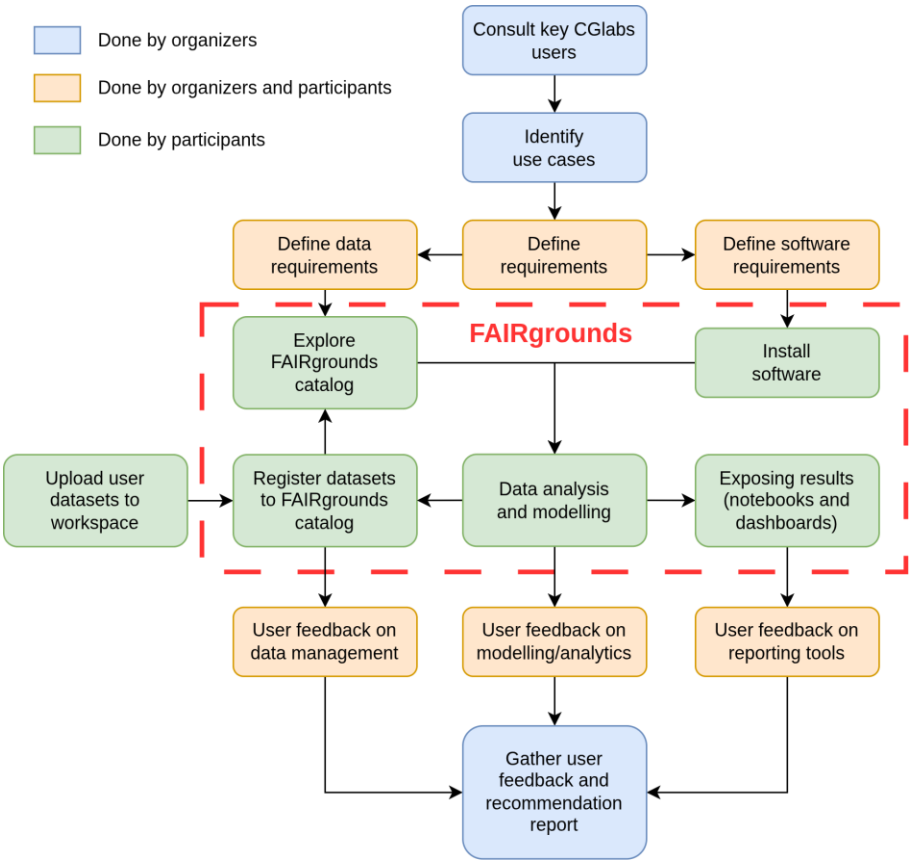


Figure 1. Proposed hackathon/workshop design

The purpose of this exercise is not to develop new innovations or full-fledged models, but rather, to test the platform end-to-end under real world conditions to observe challenges, note limitations, and register benefits of using the platforms. For this, organizers will gather feedback to be integrated into a full recommendation report for the platform. Example areas of assessment for users will include questions on data management; ease of use/configuration; and compute-store capacity. The report will also address scalability, external data sharing and platform use, and cost and sustainability concerns.

Concept note authors: Eduardo Bendito, Peter Steward, Medha Devare

ANNEX 4. USE CASE PROJECT : DOWNSCALING WAPOR DATASET

Project: Downscaling Wapor Dataset

Project goal:	To downscale WaPOR evapotranspiration from 300 m to 20 m for field-scale irrigation monitoring while preserving basin-scale consistency.

Agricultural Context

What is the agricultural (research) problem that can be tackled by solving the project goal?

The project addresses the problem of poor field-scale monitoring of agricultural water use in fragmented smallholder irrigation systems, where coarse evapotranspiration data cannot accurately capture within-field variability needed for irrigation management and water accounting.

How would the technical solution help researchers/practitioners/ growers?

It would help researchers, practitioners, and growers by providing field-scale evapotranspiration information for better irrigation monitoring, water-use assessment, and decision-making, while still remaining consistent with basin-scale. water accounting.

What would the solution enable researchers/practitioners/growers to do?

It would enable researchers, practitioners, and growers to:

- monitor evapotranspiration at field scale
- identify spatial variability in crop water use
- support irrigation management and scheduling
- improve water accounting and reporting
- assess irrigation performance in fragmented smallholder systems
- use dekadal time-series data for near-real-time monitoring, forecasting, and decision support

Technical Challenge - *What needs to be done to develop the solution?*

To develop the solution, the project needs to:

- prepare and align multi-source datasets at dekadal scale
- engineer relevant ET predictor variables
- train and optimize a machine-learning downscaling model
- validate outputs against high-resolution reference ET
- test that the downscaled product preserves large-scale consistency
- deploy the workflow as an operational tool for users

Outcome Expected - *What would the solution do to tackle the agricultural (research) problem?*

The solution would provide reliable field-scale evapotranspiration data, making it possible to monitor crop water use more accurately in fragmented smallholder systems and support better irrigation management, water accounting, forecasting, and decision-making.

Project: Downscaling Wapor Dataset

Datasets required on Fairgrounds	Links to datasets needed
List datasets needed to experiment, test and model the solution Datasets needed include: • **WaPOR Level-1 AETI** (coarse-resolution predictor ET) • **WaPOR Level-3 AETI** (high-resolution target/reference ET) • **Sentinel-2 surface reflectance** imagery • **Vegetation indices** derived from Sentinel-2, such as NDVI, NDMI, EVI, SAVI, NBR, and FVC • **CHIRPS rainfall** data • **ERA5-Land meteorological** data • **Copernicus DEM** terrain data • **ESA WorldCover** land-cover data	• **WaPOR Level-1 and Level-3 AETI:** FAO WaPOR data access and • WaPOR portal for access to WaPOR products, including current v3 access channels. • **Sentinel-2 Surface Reflectance (Level-2A):** Copernicus Sentinel-2 collection. The Level-2A product is the surface reflectance product used for deriving indices such as NDVI, NDMI, EVI, SAVI, NBR, and FVC. • **CHIRPS rainfall:** CHIRPS data download. • **ERA5-Land meteorological data:** ERA5-Land dataset. • **Copernicus DEM:** Copernicus DEM collection. • **ESA WorldCover:** ESA WorldCover data access.
Models / analytical tools required on Fairgrounds	**Links to models / tools needed**
List models / tools needed to experiment, test and model the solution Data ingestion capabilities from GCP (GEE Cloud storage) CatBoost-based ET downscaling, geospatial preprocessing, feature engineering, dekadal time-series forecasting, validation/uncertainty analysis, and dashboard visualization.	Here are the main **model / tool links**: • **CatBoost** (ET downscaling model): catboost.ai / Documentation • **Google Earth Engine** (geospatial preprocessing and raster analysis): earthengine.google.com / Documentation • **Python** (model training and workflow implementation): python.org • **Weights & Biases** (experiment tracking and model versioning): wandb.ai / Documentation • **Fairgrounds** (data sharing and analysis platform): fairgrounds.ai

ANNEX 5. USE CASE PROJECT: SPATIAL DATA CROP TYPE MAPPING

Hackathon Project Profile

Project goal: Crop type mapping in smallholder mixed farming system in Ghana

Agricultural Context

What is the agricultural problem that tackled by the project goal?

- Paucity of reliable in-season crop statistics in fragmented mixed farming systems in Ghana.
- Field measurements of crop information is labor-intensive and time consuming
- Persistent cloud cover, presence of trees and weeds complicates crop classification.

How would the technical solution help researchers/ practitioners?

An accurate map on extent of crop types will support provision of context and site-specific extension and advisory services.

What would the solution enable researchers/practitioners to do?

This model would enable:

- Identify where specific crops are grown for easier demand aggregation by traders

Technical Challenge

What needs to be done to develop the solution?

Develop an accurate (F-Score >0.85) ML model for crop type classification

Datasets required on Fairgrounds

List datasets needed to train & test the model

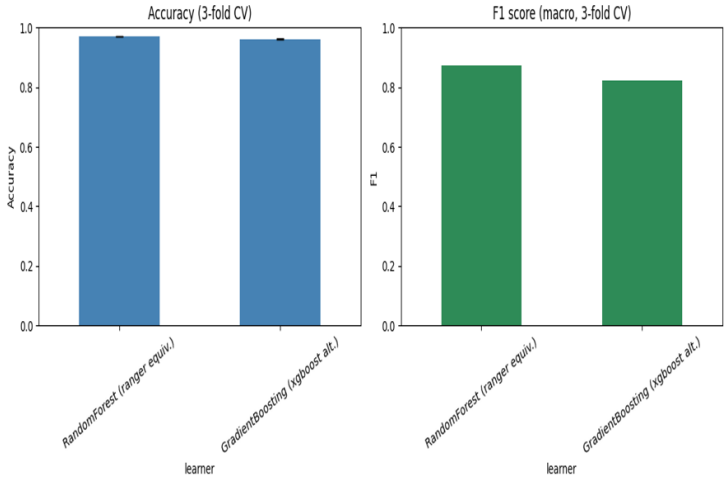
- Crop labels polygons from GoPro camera images
- Sentinel 1 & 2 imagery
- Google's AlphaEarth embeddings Imagery

Outcome Expected

What would the solution do to tackle the agricultural problem?

- Policy makers have improved accessibility and timeliness of in-season crop statistics data for food security plans

Global model accuracy assessment



ANNEX 6. USE CASE PROJECT: CROP MODEL AND YIELD PREDICTION

Project: Crop Models and Yield Prediction

Agricultural Context (*in brief*)

What is the agricultural (research) problem that can be tackled by solving the project goal?
Yield prediction remains difficult because crop performance varies widely across locations, seasons, soils, and climate conditions. Existing research is often too localized to support broader decision-making.

How would the technical solution help researchers/practitioners/ growers?
The solution would provide more consistent, spatially explicit yield predictions by combining crop, soil, and climate data. This would help researchers and practitioners better understand where and why yields differ.

What would the solution enable researchers/practitioners/growers to do? (List a key features and/or functionalities)
- Predict yields across different locations and conditions
- Compare performance under different management scenarios
- Identify areas with high yield potential or high risk
- Support better planning and targeting of interventions

Technical Challenge - *What needs to be done to develop the solution?*

- Load climate, soil, and observed yield datasets
- Run the model across different locations and conditions
- Build a workflow that can generate spatial yield predictions at scale

Outcome Expected - *What would the solution do to tackle the agricultural (research) problem?*

-Reliable, location-specific yield predictions
-Better targeting of agronomic recommendations and interventions.
-A scalable tool that can be run for other locations and seasons

Datasets required on Fairgrounds	Links to datasets needed
•SoilGrids •AgEra5 •CHIRPS	- https://isric.org/explore/ - https://www.chc.ucsb.edu/data/chirps
Models / analytical tools required on Fairgrounds	**Links to models / tools needed**
DSSAT	https://dssat.net/

ANNEX 7. USE CASE PROJECT: EMPIRICAL MACHINE LEARNING

Project: Empirical ML

Project goal: Estimate bean yield under different NPK fertilizer application rates in Rwanda

Agricultural Context

What is the agricultural (research) problem that can be tackled by solving the project goal?
In Rwanda, common bean is a key staple crop for smallholder farmers, yet yields remain low due to suboptimal and unbalanced fertilizer use. Farmers often lack site-specific guidance on appropriate NPK application rates, and crop response varies significantly across agro-ecological zones due to differences in soil fertility, rainfall, and management.

How would the technical solution help researchers/practitioners/ growers?
A predictive modelling framework that estimates bean yield under different NPK fertilizer scenarios would enable researchers and practitioners to systematically evaluate crop response to nutrient inputs and generate evidence-based fertilizer recommendations tailored to local conditions in Rwanda.

What would the solution enable researchers/practitioners/growers to do?
- *Identify optimal fertilizer rates for improved productivity*
- *Quantify the yield response to nitrogen, phosphorus, and potassium inputs*

Technical Challenge - *What needs to be done to develop the solution?*	
Develop a predictive machine learning model to simulate bean yield under varying NPK fertilizer scenarios - Ingest the cleaned trial data and environmental covariates - Develop ML models - Evaluate the performance of the models - Prediction of yield based on representative NPK scenarios	
Outcome Expected - *What would the solution do to tackle the agricultural (research) problem?*	
- A model that estimates bean yield under different NPK scenarios.	
Datasets required on Fairgrounds*	**Links to datasets needed**
- Fertilizer trial datasets which includes lat/long coordinates, NPK treatments and grain yield - Biophysical covariates:- --> Soil properties (nutrients (soil npk, texture ...)) --> Topographic characteristics (elevation, slope ...) --> Monthly growing season climate conditions (precipitation, temperature, solar radiation ...)	https://www.isda-africa.com/isdasoil/ https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1
Models / analytical tools required on Fairgrounds	**Links to models / tools needed**
- Random Forest - XGBoost	

ANNEX 8. USE CASE PROJECT: CLIMATE MODELING

Project: Translation and localization (downscaling) cross time-scale climate forecast

Project goal: Translate seasonal climate forecasts into climate intelligence/advisory that enables governments, extension systems, and farmers to manage crop and livestock risks.

Agricultural Context (*in brief*)

Agricultural Problem Farmers need climate info at weekly/monthly scales to manage planting, irrigation, and livestock—but forecasts only give 3-month averages. This mismatch leads to mistimed decisions, wasted inputs, and vulnerability to droughts, floods, and heat stress.

Technical Solution Transform seasonal forecasts into monthly risk intelligence: precipitation anomalies, rainfall onset timing, drought/wetness indices (SPEI), and livestock heat stress maps. Analog-based methods link forecasts to historical conditions for better interpretation. Delivered through digital platforms as actionable, localized advisories.

What It Enables Farmers and advisors can time planting to avoid dry spells, select drought-tolerant varieties, optimize irrigation, manage pests proactively, and reduce livestock heat exposure. Shifts decision-making from reactive to anticipatory—improving resilience and efficiency.

Technical Challenge - *What needs to be done to develop the solution?*

- Establish automated data pipelines with national meteorological services and global providers to ingest daily and monthly climate data, including seasonal forecasts, historical observations, and reanalysis datasets. This requires harmonization, bias correction, spatial downscaling, and quality control to ensure physically consistent and operationally reliable inputs.
- Develop an integrated climate analytics framework that combines precipitation anomaly forecasting, onset-of-rainfall detection using agronomic thresholds, and drought/wetness classification based on SPEI. An analog-based matching approach will link forecasts to similar historical conditions, improving interpretability and enabling probabilistic risk scenarios.
- Automate monthly map generation and coordinate with partner institutions to integrate outputs into existing agricultural advisory systems (extension bulletins, SMS alerts, dashboards).

Outcome Expected - *What would the solution do to tackle the agricultural (research) problem?*

The solution delivers monthly maps showing drought or excess moisture risk by location. Farmers use these to make concrete decisions: in "Very Dry" areas, postpone planting, select drought-tolerant varieties, and activate irrigation; in "Very Wet" zones, prepare drainage and monitor for disease. Extension services distribute maps through existing channels (SMS, radio, WhatsApp), translating forecasts into field-level actions. By grounding predictions in historical analogs ("conditions like 2015"), the maps build trust and reduce climate-related losses while optimizing input costs.

Datasets required on Fairgrounds	Links to datasets needed
• Dailly ECMWF SEAS5 daily Hindcast and forecast data • Monthly gridded precipitation (1982–2025, CHIRPS or equivalent, ~5 km resolution) • Monthly potential evapotranspiration (ERA5-Land or TerraClimate) • Seasonal forecast ensembles (NMME, IRI, C3S, or regional products like IDEAM CPT) • Administrative boundaries for spatial subsetting • Crop calendar datasets to align forecast windows with phenological stages	**CHIRPS:** chc.ucsb.edu/data/chirps **IDEAM:** bart.ideam.gov.co/wrfideam/new_modelo/CPT/ **Natural Earth:** naturalearthdata.com/downloads/ **SPEI-3:** Internal CGIAR (SAT_PMA) CMIP6 data **ECMWF SEAS5:** hindcast and forecast dta
Models / analytical tools required on Fairgrounds	**Links to models / tools needed**
Python 3.9+ with xarray, rioxarray, rasterio, geopandas, matplotlib, netCDF4, requests, pyproj, Jupyter Notebook environment. Climate forecast model (NexGen approach)	https://dbc-5db1353f-aaf4.cloud.databricks.com/editor/notebooks/4251821357465766?o=7474657369545197#command/5671858097073032 https://iri-pycpt.github.io/PyCPT2-Seasonal-Forecast-User-Guide/cpt.html

ANNEX 9. IN-WORKSHOP FEEDBACK SURVEY RESULTS

1. How easily were you able to find the data you need in Fairgrounds?

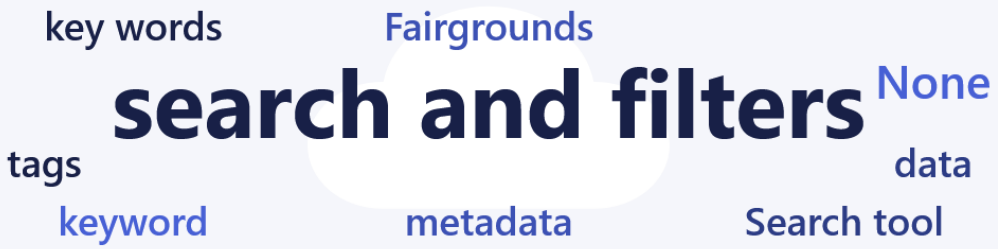
2. Were you able to:

3. What tools or features helped you discover data in Fairgrounds? (e.g., search, filters, metadata, tags)

18
Réponses

"Search"
"Search tool"
"I didn't find any of the data that I need in Fairgrounds."
...

4 répondants (22%) répondu search and filters pour cette question.



4. What challenges have you encountered when searching for data in Fairgrounds?

18
Réponses

"NA"
"Data was referenced but difficult to integrate in an existing pipeline"
"There is no structure to the data to explore what is present under different t... "
...

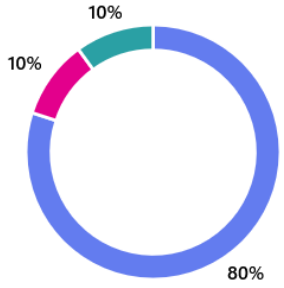
2 répondants (11%) répondu topic pour cette question.

5. Have you searched for the same data in GARDIAN ?

No	8
Yes	10

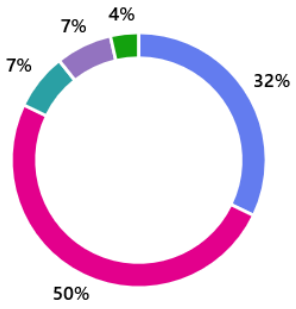
6. How was it to find the data you need in GARDIAN?

Easy	8
Neutral	1
Difficult	1
NA	0



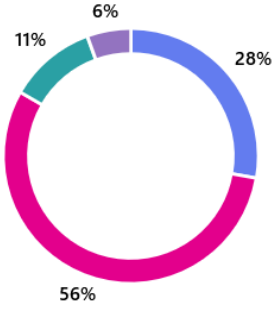
7. Which data types have you added to the catalog of Fairgrounds?

Survey/trial data	9
Spatial/GIS data	14
Images	2
None	2
Other	1



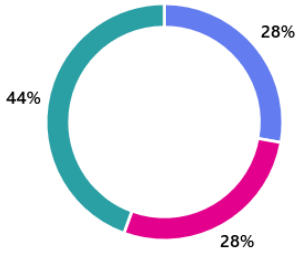
8. How would you rate pulling data into workflows in Fairgrounds?

Easy	5
Neutral	10
Difficult	2
NA	1



9. For the data that you are using, if you were a naive user would you be able to understand enough about the data to c onfidently use it from the metadata template Fairgrounds is currently proposing?

Yes	5
No	5
NA	8



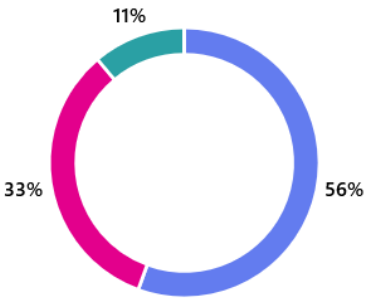
10. Could you elaborate and suggest improvements?

4
Réponses

"An improvement could be to add a tag to the data that is already available o..."
"Some thoughts: https://github.com/CGIAR-Climate-Data-Hub/metadata/blo... "

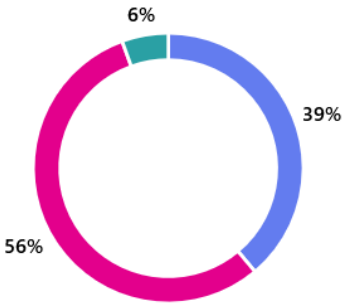
11. Are data rights, permissions, and license clear and understandable?

Yes	10
No	6
NA	2



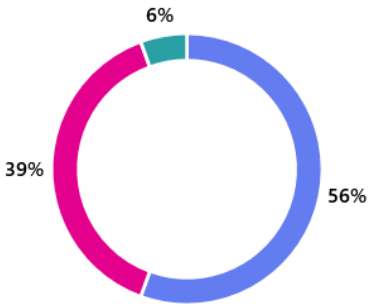
12. Have access restrictions prevented you from using data effectively?

Yes	7
No	10
NA	1



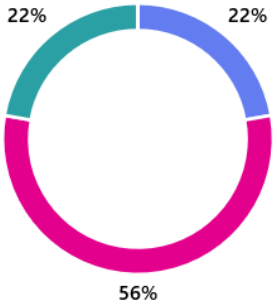
13. Is the authentication process (login, roles, access levels) clear and easy to use?

Yes	10
No	7
NA	1



14. Do you have concerns about data security or privacy?

Yes	4
No	10
NA	4



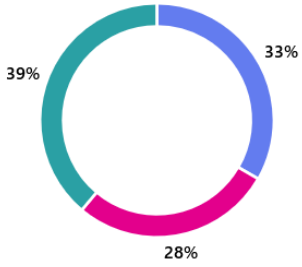
15. Please describe yours concerns

4
Réponses

"What are security policies when it comes to private datasets? How can the d... "

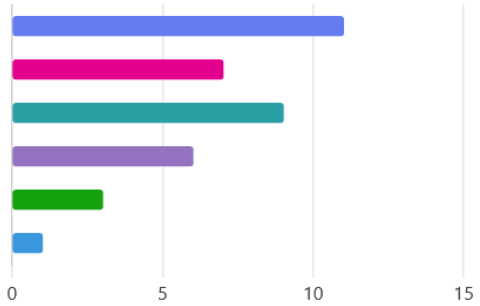
16. Is the data and governance-related guidance/documentation available on Fairgrounds sufficient?

Yes	6
No	5
NA	7



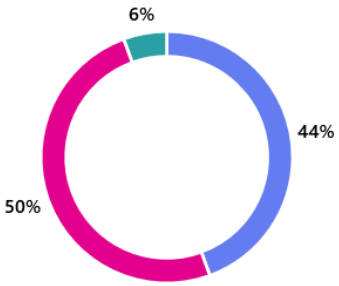
17. What data governance considerations would you like to see improved in Fairgrounds?

Access	11
Ownership	7
Citation	9
Stewardship/roles	6
Ethics	3
Other	1



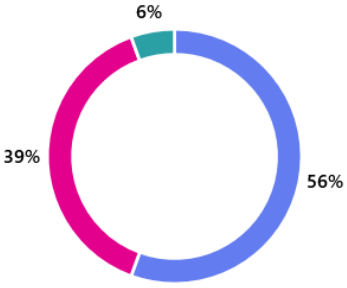
18. Were the software and libraries needed to conduct your analysis readily available (by default or in the custom computing environments available)?

Yes	8
No	9
NA	1



19. There is some AI integration into Databricks; was it useful?

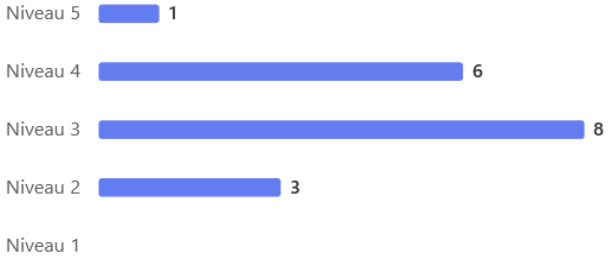
Yes	10
No	7
NA	1



20. How do you rate the native code editor in Databricks?

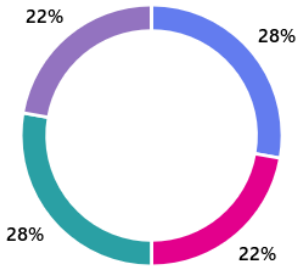
3.28

moyenne



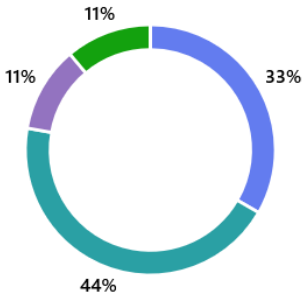
21. If you wanted to use an integrated development environments (IDE) like VSCode or RStudio, were you able to?

Yes	5
Partially	4
No	5
NA	4



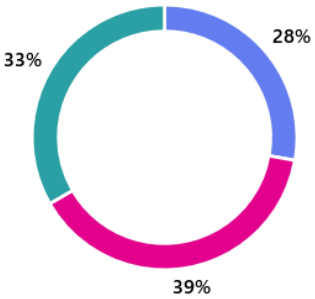
22. Which tools/models have you attempted to install?

Machine Learning models	9
Other AI	0
Geospatial	12
Crop models	3
Other	3



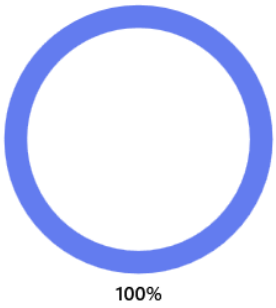
23. How would you rate installing analytics tools/packages/libraries?

Easy	5
Neutral	7
Difficult	6



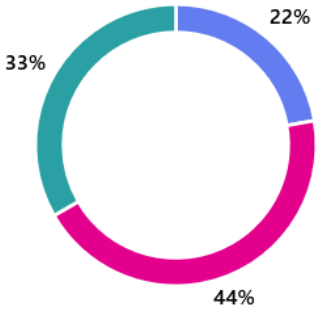
24. How important is it for you to be able to install analytical tools/packages/libraries yourself?

Extremely important	18
Not very important	0
Prefer tech team to do this for me	0



25. Was the guidance/documentation available sufficient to enable you to configure analytics needs easily?

Yes	4
No	8
NA	6

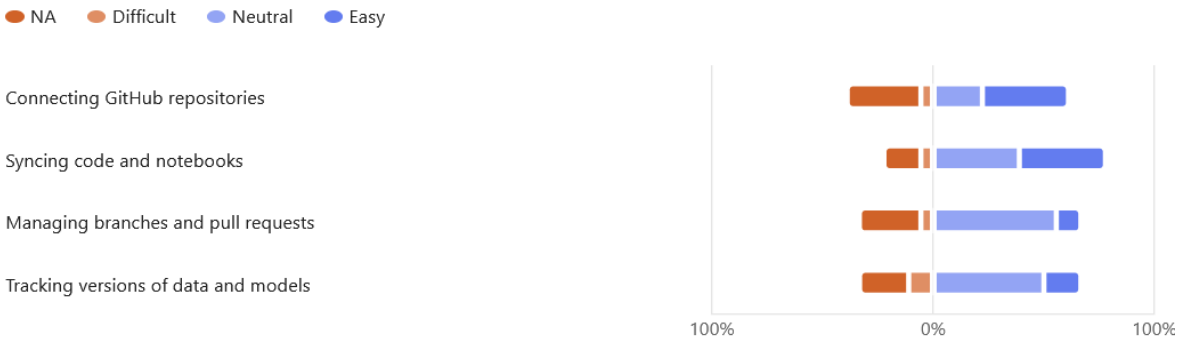


26. What additional documentation or training would improve your experience?

8 Réponses

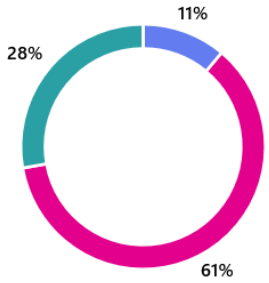
ID ↑	Nom	Réponses
1	anonymous	User manual
2	anonymous	Data sets linkages
3	anonymous	How to setup and optimize the compute resources
4	anonymous	Short videos. The available documentation and training are quite scattered
5	anonymous	user manual, and use case examples
6	anonymous	Documentation on using R would be valuable. Most of the documentation currently is for Python users.
7	anonymous	Tier level explanation, output registration, search for other's datasets
8	anonymous	I think there are so many links to supporting tutorials that it will take more time to go through all of them. Also, there is a need to cater for versioning as the versions transform

27. What was your experience in terms of



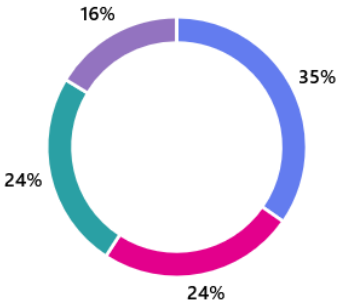
28. From your perspective, what level of technical knowledge is required to use the platform's analytics capabilities?

Beginner (minimal coding)	2
Intermediate (basic scripting)	11
Advanced (DevOps, cloud, ML engineering)	5



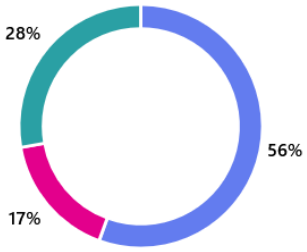
29. What technical stack knowledge is required?

Python	17
R	12
Cloud infrastructure	12
APIs	8
Other	0

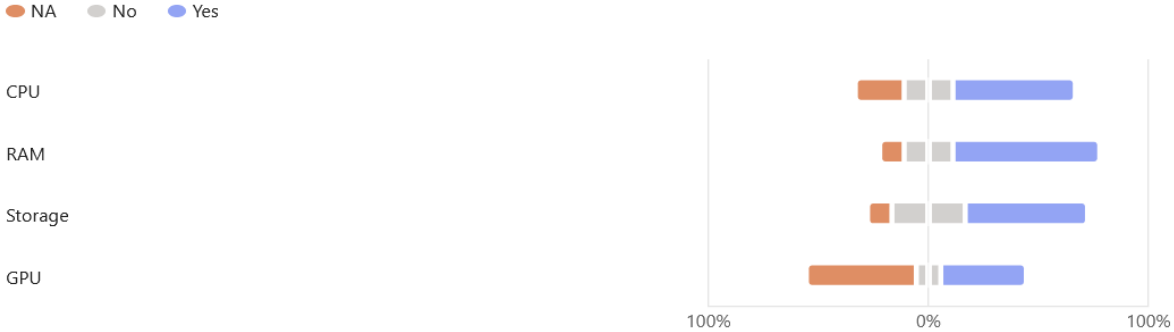


30. Did you experience performance limitations when working with large datasets?

Yes	10
No	3
NA	5



31. Do the compute-store resources available meet your needs?



32. Were there specific configurations or hardware that you would like to have available?

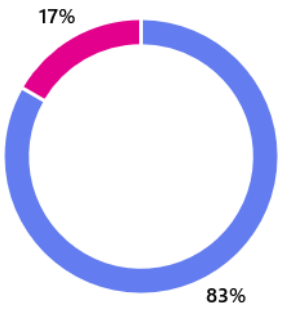
33. Tell us about specific hardware or features you would like to see.

7 Réponses

ID ↑	Nom	Réponses
1	anonymous	etw
2	anonymous	NA
3	anonymous	Job Scheduling & Managment, enviroment creation
4	anonymous	Terra. Mostly R packages to be more accessible.
5	anonymous	Currently it is so difficult to work with the R cluster without some hustle. I use bith R and Python. So this platform should allow working with R environment semless with a good management of the librairies.
6	anonymous	Spatial set-up for python geospatial processing seems good.
7	anonymous	Serverless computing cluster with GPUs Costs dashboards associated with my user sessions

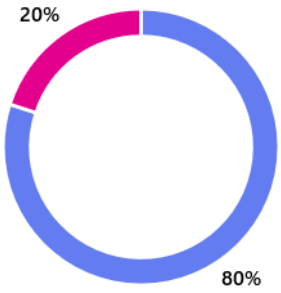
34. Do you use CG Labs?

Yes	15
No	3



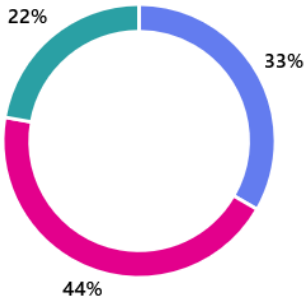
35. Does CG Labs meet your computing needs?

Yes	12
No	3



37. Is the available documentation on Databricks sufficient?

Yes	6
No	8
NA	4



38. What additional guidance, tutorials, or examples would be helpful?

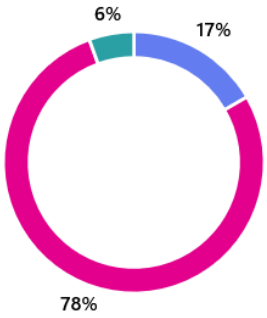
8 Réponses

ID ↑	Nom	Réponses
1	anonymous	Simplified video tutorials.
2	anonymous	Basic functionalities
3	anonymous	It's hard to get to the specific tutorials, and the example isn't very useful for our case either. Most of the tutorials don't focus on a specific area, so it needs a lot of work to get better.
4	anonymous	Setup and optimizing compute environment, troubleshooting the installation of libraries
5	anonymous	short videos rather than long articles to read
6	anonymous	R configuration, common line interface, data ingestion, data outputs sharing, satellite imagery processing
7	anonymous	Documentation could be added as a tab on databricks menu to facilitate access
8	anonymous	Training and documentations

39.What are your expectations regarding tech support as you consider further use of Fairgrounds?
lots
NA
Not sure.
None
tech support
None
Most of the backend needs an expert to tech-assist the domain-specific experts. Otherwise it need institutional focal experts who can support across the board for all users or developers
Continous and dedicated support will be available
easy communication and response tech team
technical support is very good. but still, we need to explore for real time use cases
Chatbot for FAQs.
support about databricks resources, dataset ingestion
The tutorials must provide some examples as error diagnosis was challenging
A technical backstopping when exploring the full potential of Fairgrounds and implementing real usecases
It is overwelming. It could be good if thos could be a bit straight to the point.
Setting up compute.
Administration of access and use for teams (e.g. setting projects having use-limits for the project teams).
Trouble shooting.
Some tasks are not easy to complete without expert guidance; therefore, continuous support and training options are expected going forward
Yes

40. How would you rate finding your way on Fairgrounds?

Easy	3
Neutral	14
Difficult	1



41. Overall, how would you rate Fairgrounds?

3.39
Average rating

42. Provide any further comments or thoughts that you would like here.
We couldn't try fairgrounds since the only tool available for our use cases was the databricks. So it's diffcult to evaluate fairgrounds. There is a lot work done with other initiatives that should be considered or at least integrated such as GARDIAN, Carob...
Connect with local information or connect with column data that i can use easy
I need to explore more
The system need rigourous tests to stabilize the access to compute and asset resources
facing difficulty in configuration, settings and path. we are facing issue in large data analysis.
Provide some connection between Databricks and fairgrounds. Make the documentation easily accessible. Improve on the permission issues when working with R scripts. Python is easily accessible, so kudos for that bit.
Our main interaction has been through Databricks, as the utility of Fairgrounds is still not entirely clear. Moving forward, I would expect the platform to have better development and a more intuitive interface. Ideally, it should allow publicly available datasets to be linked directly to Databricks.
There is a need to provide additional support to remotely sensed data and Google Earth Engine to cater for the slow processing in Databricks
For now it is hard to really see what FairGround is capable of or not. We mainly use Databricks. Is FairGround going to enable all the features of Databricks? I doubt so it could be good to know the boundaries (maybe showing the roadmap). In the current version, I ony upload data to Fairground. I was expecting that the workspace in Fairground works and team members can collaboratively develop a solution. It is good that we could test some pipeline from the Fairground but it should go beyond that.
The storage element of FAIRgrounds is very useful, we will consider using it for the AAA Atlas and Climate Data Hub. However, our data has a specific structure (e.g. atlas_data/exposure/mapspam2020) that links to a STAC catalog are we able to maintain the hierarchy of the data in the TierIII FAIRgrounds? or is it flat? Spatial users may have large storage requirements 10s, 100s of GB. There are many foundational datasets for our geospatial work that are not currently available in DataBricks or FAIRgrounds. If we are to work in this environment we will need them. Some may be present in OpenAWS, many are present in GoogleEarthEngine, some we may need to host. Is there a way to clone datasets from GEE into Databricks?
The integration of Databricks and Fairgrounds is confusing and could be better explained
It is an excellent platform to work on

ANNEX 10. USE CASE GROUP DISCUSSION FEEDBACK

Climate Use Case feedback

1. What improvements do you expect or desire (compared to current state) for future development of Fairgrounds, considering the following:

- Data ingestion, registration, access
  Integrations or automation pipelines that connect with data sources (OneDrive) that instead of copying files in Databricks can read directly (e.g. GARDIAN)

- What could/should be done faster / more easily
  We couldn't use the data ingestion option in fairgrounds, it's not intuitive or documented. From FG, users are brought to Databricks without explanation and get lost.

- Anything else?
  - It would be helpful if there are use cases implemented with a complete pipeline, from data ingestion until results before the workshops.
  - More time to test the platform, it is difficult to work on the use cases asynchrony with day-to-day work.
  - How is fairgrounds related to databricks (why is this the best solution), is there any documentation available?

2. What do you suggest to reduce effort in managing data in Fairgrounds?
   Template notebooks + cluster configs bundled with each use case so new analysts reproduce a run in one click.

3. What features / capabilities / benefits do you want to see that are currently missing on the platform?
   Native dataset versioning (snapshot + diff) so teams can track how input files evolve over time. A data lineage view (visualize how data go from raw → tier0 → tier1 → tier2)— visualized automatically from existing Unity catalog metadata.

4. What do you see as the top 2-3 barriers or concerns to using Fairgrounds?
   - Serverless compute breaks geospatial libraries (rasterio, netCDF4) — forces every team to manage custom cluster configs
   - Connection with local storage (on-premises servers)
   - Price, sustainability and team adoption
   - Need more documentation
   - How to track users/team usage over time, and track pricing

5. In general:
   - Astonished, excited by the Databricks platform. It has so many capabilities, but it’s a big overwhelming for new users.

- The platform seems to have been built for private sector and is not optimized for CGIAR so far.

Spatial Data Use Case feedback

1. What improvements do you expect or desire (compared to current state) for future development of Fairgrounds, considering the following:
   - Data ingestion, registration, access: easy access to code/scripts (right now we have to look for it)- maybe on the left panel.
   - Easily accessible information/guides on how to migrate data from other cloud storage/databases
2. What do you suggest to reduce effort in managing data in Fairgrounds?
   - Understanding on creating a compute for better efficiency. Users want to know the computing needed for the work they plan to run.
3. What features / capabilities / benefits do you want to see that are currently missing in the platform?
   - Interoperability amongst Fairgrounds and GARDIAN, Community Forum where people discuss issues
4. How does fairgrounds compare w/ existing approaches
   - Would be nice to have the integration of ML ready tools because it's useful.
   - Scalability of compute clusters to facilitate big data processing
5. What do you see as the top 2-3 barriers or concerns to using Fairgrounds?
   - **Training:** We need to get trained on Fairgrounds/ Data Bricks/ Gardian/Carob. Not everyone was on the same page at the start of the workshop - would be nice to know more before doing the feedback survey.
   - Have some FAQs. Use case example – hands on experience.
   - Developed applications (e.g. dashboard) can't be shared in public without credentials

Crop Model – Yield Prediction Use Case feedback

1. What improvements do you expect or desire (compared to current state) for future development of Fairgrounds, considering the following:
   - Data ingestion, registration, access
     - Unclear data ingestion processes and procedures, not enough documentation
     - Need to use the notebook for data ingestion. But this is for programmers - not easy for normal data scientists.
     - Tier/folder structure was not explained before.

- What could/should be done faster / more easily: Moving files between folders.
- Anything else: Lack of clarity on the creation of internal computing, as the workflows broke due to space limitations.

- Model installation
  - PyDSSAT needs temp folder for installation of DSSAT. It could be good to have DSSAT installed.
  - Need examples of other models.

- Analytical workflow implementation
  - Automatic language translation e.g. from Python to R
  - Add R Studio

- Ability to collaborate
  - None – This is straight forward.
  - We need versioning ability of the scripts, just like in Git hub.

- Sufficiently powerful, responsive computing
  - Termination of simulations due to limitation in computing resources. Setting up of the computing environment should be simpler.
  - Add more cores (need more computing power)

2. What do you suggest to reduce effort in managing data in Fairgrounds?
   - Create a simpler interface for data registration, search and ingestion
   - Data version tracking e.g. modified data (need for documentation)
   - List of datasets downloaded into Fairgrounds to reduce redundancy

3. What features / capabilities / benefits do you want to see that are currently missing in the platform?
   - Auto programming language translation from Python to R and vice versa
   - Better AI code enhancement efficiency
   - Add R studio

4. How does Fairgrounds compare w/ existing approaches?
   - CG Labs simpler and straightforward interface
   - Fairgrounds has integrated data and analytics capability, CGLabs has but loosely connected to GARDIAN.

5. What do you see as the top 2-3 barriers or concerns to using Fairgrounds?
    - Limited documentation on how to use Fairgrounds.
    - Limited information on what kind of tools are available e.g. We just found out now that there is Roboflow.
        - No R Studio
        - Sharing outputs in an interactive way non platform users.

Overall, Fairgrounds has great potential for data management.

Crop Model – Empirical ML use case

1. What improvements do you expect or desire (compared to current state) for future development of Fairgrounds, considering the following:
    - Data ingestion, registration, access
        - Faster upload, especially for large files (took over 16h sometimes)
        - Increase the upload of storage files for “small files," like 10GB. Currently need to contact the developer for files larger than 5GB.
    - What could/should be done faster / more easily: Upload of files
    - Anything else
        - APIs with existing data repositories, such as Carob
        - Metadata when registering data was a bit much.
    - Model installation : Geodata support
    - Analytical workflow implementation
        - Running models, in the serverless compute
        - Permission management of packages/clusters
    - Ability to collaborate : It was fairly easy.
    - Sufficiently powerful, responsive computing : A bit slow on serverless computing when running ML models
2. What do you suggest to reduce effort in managing data in Fairgrounds?
    - Less reliance on developers, e.g., having to contact them to upload large files.
    - Improve the file structure for the data. It currently looks like a database schema.
3. What features / capabilities / benefits do you want to see that are currently missing in the platform?
    - UI/UX to be improved to easily find your way around the platform as a new user.

- Make it easy to manage the data writing files for some packages (Geodata, for instance)
- How to connect with local IDEs?

4. What do you see as the top 2-3 barriers or concerns to using Fairgrounds?
   - Too developer-reliant, e.g. contacting them for uploading large data.
   - Supported more languages like R better. Make it easier to set up R. Most ag researcher work with R.
   - Smoother connection between DataBricks and Fairgrounds.
5. How does Fairgrounds compare w/ existing approaches?
   Pros:
   - One-stop shop for data storage + computing
   - Compute size is good
   - Pre-installed packages for Python

   Cons: Difficult user experience. Not very intuitive.

ANNEX 11. PRE AND POST WORKSHOP ONLINE SURVEYS

Pre-workshop survey results: Pre-Hackathon Survey_04-2026.xlsx

Post-workshop survey results: Post-Hackathon Survey_04-2026.xlsx

Report written by Céline Aubert, Medha Devare
July 2026
Fairgrounds Platform Hackathon - Report

Citation : C.Aubert, M.Devare, 2026. Fairgrounds Platform Hackathon. Workshop report, IITA, 20p.

IITA
Transforming African Agriculture

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
IFPRI