Data Harmonization Workshop
Towards FAIR and AI-Ready Data

Summary Report

- Workshop conducted from June 17th to June 20th, 2025, as part of AoW1 of CGIAR's Digital Transformation Accelerator
- This report is being shared on July 7th, 2025

Sections

1. About the Workshop (p3)
2. Objectives, Context, and Approach (p6)
3. Outcomes, Outputs, and Next Steps (p12)
- Appendix (p22)

CGIAR

DIGITAL TRANSFORMATION

Executive Summary

Workshop Premise
The Data Harmonization Workshop brought together members of 12 CGIAR Centers in a hybrid four day event to **develop a practical approach to foster a FAIR and AI-Ready data ecosystem** across Centers, starting with data publishing.

Building on past efforts, the workshop prioritized consensus building, **focusing on key agreements for data publishing** in two initial domains: Agronomy and Socioeconomics & Gender (with planned inclusion of further domains).

Workshop Achievements
Outcomes include **Key Agreements** to steer further iteration and implementation of Data Harmonization efforts for AoW1 of the CGIAR Digital Transformation Accelerator.
/ Skip directly to **Key Agreements** (p13)

Outputs included (1) **Draft Harmonization Guidelines**, (2) **Core Variables** for agronomy and socioeconomics, and a (3) **Two-pathway Model for Adoption** (Mandate and User Motivation).

Next steps include finalizing v0.1 of Harmonization Guidelines, expanding coverage across additional domains, and initiating the work plan agreed during the session.

1 —
About the Workshop





Workshop Overview

Goal

Enable harmonization of data* across CGIAR Centers and Programs

*broadly to include practices, processes, protocols, systems, but starting with data publishing

Format

- Online Demo Day: June 16th
- Workshop Days: June 17th to June 20th
  - In person at CIMMYT HQ in Texcoco, Mexico
  - Virtual asynchronous participation by attendees across Centers
- Agenda shown in the Appendix

Participation

In-person participants
worked collaboratively during the day (9am to 5pm Mexico City Time); integrated feedback from virtual participants.

Virtual participants
contributed to shared documents asynchronously, adding to in-person inputs from each previous day; participated in daily moderated synchronization calls for alignment.

[Full participant list included in Appendix]

Centers represented in-person
Alliance of Bioversity and CIAT, **CIMMYT**, **IFPRI**, **IITA**

Centers represented remotely
Africa Rice, **CIFOR-ICRAF**, **CIP**, **ICARDA**, **ICRISAT**, **IITA**, **ILRI**, **IRRI**, **IWMI**, **WorldFish**

2 — Objectives, Context, and Approach

Workshop Objectives

- **Create a framework** (or mental model) for data harmonization 
- **Agree on the framework** for cross-center data harmonization (acknowledging trade-offs; getting as close to consensus as feasible in the available time) 
- **Define and commit to a workplan** to validate and implement the agreed data harmonization framework 

Strategic Context

- **Increased donor pressure** for data sharing across CGIAR Centers
- **Spotlight on AI** demands focus on everything surrounding data
- **Shift in narrative from standardization to harmonization** to drive adoption
  - Common guidelines, not imposed standards
  - Minimum viable agreement not as a limitation, but as a baseline to build upon
- **Commitment to a participatory process** required to consider diverse perspectives on data and to build upon prior efforts
  - Acknowledging strengths and limitations of existing data policies and workflows
  - Sharing best practices
  - Prioritizing practicality and iteration
  - Avoiding added burden on resources
- **Future Data and Legacy Data are two *Flagship* areas in AoW1** of CGIAR's Digital Transformation Accelerator

Workshop Components

Showcase of methods / tools / approaches

- Demo Day presentations by representatives from CGIAR Centers
  - Folder with presentations
- FAIR Data by CABI
- Ontologies and AI Readiness by Marie Angélique Laporte (Alliance

Review of existing frameworks and resources to be leveraged

- Carob / Data Standardization Guidelines and Scripts
- CABI / FAIR Self Assessment tool
- RHoMIS / Rural Household Multi-Indicator Survey

Mixed dynamics for evaluative and generative activities

- Open discussion with prepared prompts
- Guided mapping exercises (processes, audiences, tasks)
- Group breakout sessions

Workshop Working Principles

- **Leverage what already exists**; don’t start from zero
- **Prioritize actionability over completeness**; minimum viable outputs become primitive building blocks; avoid attempting to resolve too much at the outset
- **Validate ideas developed in the workshop after the workshop** with users (who will be following Data Harmonization Guidelines) and experts (to help inform iteration of frameworks and guidelines)

Domains covered

To ground the drafting of Data *Core Variables* and a Data Harmonization Guidelines document, two Domains were prioritized based on domain affinity with in-person participants.

Additional Domains will be covered following the Workshop.

Domain
Agronomy

Source of Initial set of variables brought into the workshop to iterate upon:

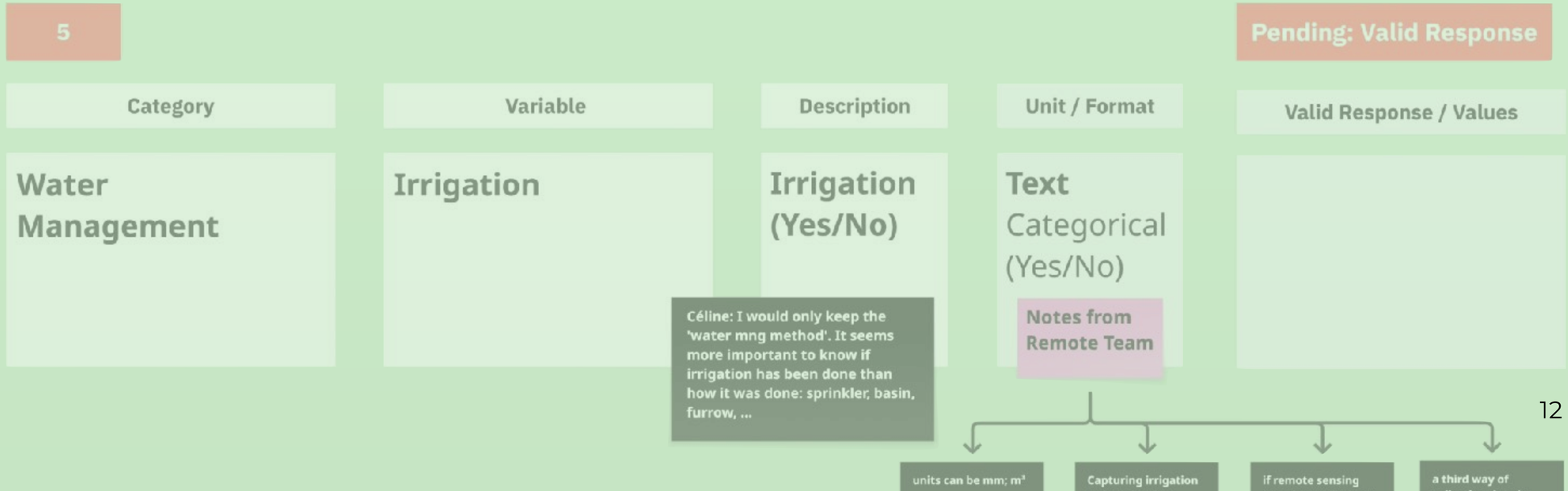
- **Excellence In Agronomy Initiative** (A joint effort between almost of all CGIAR Centers)

Domain
Socioeconomics & Gender

Source of Initial set of variables brought into the woekshop to iterate upon:

- **Big Data Platform** (Cross-Center effort)

3 — Outcomes, Outputs, and Next Steps



Outcome: Key Agreements made during the workshop

1
The Initial focus will be on data publishing, not data collection.

2
A *Guidelines Document*, then a *Pitch Document* should be the first two outputs to drive adoption of data harmonization across CGIAR Centers.

3
Consider two adoption pathways: *Mandate-driven* (top-down) and ***Motivation-driven*** (bottom up) to drive data harmonization.

4
A *Guidelines Document* will include additional recommendations beyond *Core Variables*: naming, metadata, attribution, other conventions.

5
***Core Variables* are not intended as a fixed ceiling** – they are just the baseline; *Extended Variables* can be defined to complement *Core Variables*.

6
A *Guidelines Document* is intended primarily to guide future data publishing; this does not, however, exclude use for evaluation and cleanup of historical data.

7
Workshop participants will be Champions for data harmonization at their Centers.

8
Workshop participants are committing to a set of tasks identified and assigned during the workshop.

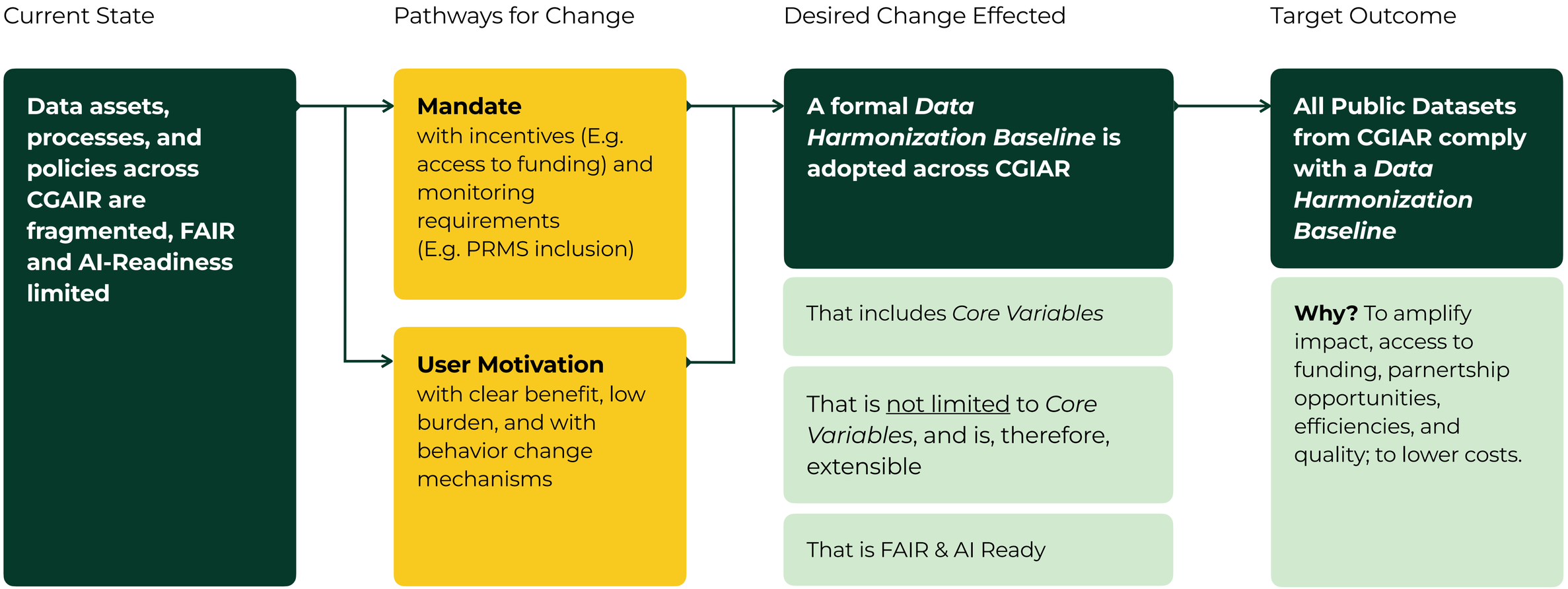
Pending Further Alignment

1
Stick with Domains as the entry point for *Core Variables?* Or again consider unit of observation/ analysis?

2
Should a “Global List” of variables shared across domains be defined? How?

3
How and when should Ontologies be established?

Schematic Summary
showing aspiration



Outcome: Audiences Identified

Initial Target Audiences

- **CGIAR Scientists** (including Data Scientists)
- **Senior Managers** at CGIAR Centers
- **Senior Managers** at CGIAR Science Programs

Additional Audiences

- **CGIAR Leadership**
- **Funders**
- **Donor Funding Unit**
- **Research Partners**
- **Ethics Committee**
- **Data Collaborative** (group to be formed)

First step

For any audience:
Make the case to get *buy-in*
with clear value proposition for Mandate sponsors

and

clear incentives for adopters

Target Action

- **Follow Data Harmonization Guidelines Baseline**
- **Institute & deliver a Data Harmonization Mandate**

Output: Documents
started during the workshop

Draft / Word Doc
Data Harmonization Guidelines

Plan / Word Doc
Starter Task List
with assigned owners

Draft / Miro Board
Core Variables
Agronomy Domain

Reference / Word Doc
Index of Resources & Examples
Guidelines, Standards, Ontologies (including FAIR cookbook, TermInAg,, Carob, SDC handbook, Excellence in Agronomy, Gardian, RHoMIS, CG Core Metadata, more)

Draft / Miro Board
Core Variables
Socioeconomics & Gender Domain

Next Steps

- **Finalize Harmonization Guidelines v0**

- **Iterate Core Variables for initial Domains**
  - Agronomy
  - Socioeconomics & Gender

- **Conduct Core Variables exercise for additional Domains**
  - Crop Breeding
  - Livestock (including Aquatics)
  - Environment & Climate

- Initiate several required processes: **Planning**, **Validation Strategy**, **Communications Strategy**, **etc.**

The task list created during the Workshop provides more detail on the items listed here and additional activities (not listed here) that will follow the Data Harmonization efforts after the Workshop.

Task List / Prioritized items

Note: These and additional tasks will be managed via a Project Management Tool to be set up as part of Task 4 in this table:

	Task	Owner(s)
1	Package Workshop Report (Note: this document is the Report)	**ROBERTO**
2	Seek Endorsement from Workshop Participants' Leaders	**ANDREA**
3	Complete Core Variables for: Agronomy, Socioeconomics & Gender, Crop Breeding	**KATE, MEDHA, CARLO**
4	Set up Management Plan, Tools, and Processes for Data Harmonization Tasks	**ANDREA, DAVID**
5	Establish Data Harmonization Champions / Reps for each Center	**ANDREA**
6	Complete v0 of Data Harmonization Guidelines Document	**ANDREA, MEDHA, (+Delegated Sections)**
7	Define Consultation Plan for Key Factors: Privacy, Infrastructure, Data Ecosystem, etc.	**ANDREA, MEDHA, MARIE ANGÉLIQUE**
8	Package Workshop Outputs (Minimum Variables, Guidelines Document)	**SATISH, CÉLINE, KATE**
9	Task Core Variables for Additional Domains: Livestock, Environment & Climate	**ANDREA, KAI**
10	Design a Communications and Engagement Strategy	**SATISH, CÉLINE**

A Call to Action

Be a champion for Data Harmonization in your CGIAR Center and the Program(s) you are involved in.

★

Start by sharing this document with colleagues, visit the linked resources, and join future Data Harmonization Sessions.

Contact Andrea Gardeazabal a.gardeazabal@cgiar.org to be added to the Teams Channel and Invite List.

Start, advance and complete your assigned tasks.

★

Look for your assigned tasks earlier in this docuement on p19; also in the linked documents on p17; and also to be shared via a forthcoming project management tool.

Give feedback, comment, and propose.

★

Visiting the linked resources (Word Documents, Miro Boards, external links, on p17), provide feedback on them, and add your contributions directly on them, with your name so we can trace them to you.

Thank you

CGIAR

DIGITAL TRANSFORMATION

Appendix

Workshop Participant List

In Person Participants
Azzarri, Carlo (IFPRI)
Devare, Medha (IITA, SO)
Dreher, Kate (CIMMYT)
Gardeazabal, Andrea (CIMMYT)
Laporte, Marie Angelique (Alliance)
Nagaraji, Satish (CIMMYT)
Sonder, Kai (CIMMYT)

In Person Drop-in Guests
Fonteyne, Simon (CIMMYT)
Moretto, Andre (CIMMYT)
Snapp, Sieglinde (CIMMYT)

Facilitation, Organization, Capture
Christen, Roberto (External)
Garcia, David (CIMMYT)
Gardeazabal, Andrea (CIMMYT)
Nuñez, Daniel (CIMMYT)

Remote Participants
Ali, Ibrahim (AfricaRice)
Anilkumar Vemula (ICRISAT)
Atassi, Layal (ICARDA)
Attaher, Samar (ICARDA)
Aubert, Celine (IITA)
Bartolini, Pietro (ICARDA)
Bendito, Eduardo Garcia (IITA)
De Leon, Dehner (IRRI)
Domelevo Entfellner, Jean-Baka (ILRI)
Erlita, Sufiet (CIFOR-ICRAF)
Gakhar, Shalini (IRRI)
Ghosh, Surajit (IWMI)
Ismail Mohammed (ICRISAT)
Juarez, Henry (CIP)
Kouadio, Amani Louis (AfricaRice)
Kiala, Zolo (IWMI)
Lecoutere, Els (ILRI)
Longobardi, Lorenzo (WorldFish)
Muchiri, Caroline (ILRI)
Mudereri, Bester (CIP)
Murali Kr Gumma (ICRISAT)
Niyati Singaraju (IRRI)
Poole, Elizabeth Jane (ILRI)
Radanielson, Ando (IRRI)

Links to Transcripts

Day 1
June 17th

- AI Summary
- Folder with Additional Notes

Day 2
June 18th

- AI Summary
- Folder with Additional Notes

Day 3
June 19th

- AI Summary
- Folder with Additional Notes

Day 4
June 20th

- AI Summary
- Folder with Additional Notes

Reasons for limited success of past data initiatives across CGIAR

Crowdsourced answers during the workshop

Lack of incentives; lack of ownership	Data systems not ready / technical limitation	Lack of awareness	"What's in it for me"? Mindset (Scientist POV)	Not co-created; Not just build and validate, but co-create from the start	Other roles needed? (E.g. Data Managers) (to alleviate burden on scientists)	No incentive to publish data	Do we need dedicated staff to focus on Data workflows?
Culture factors (what is expected of researchers, and what isn't)	Need for dedicated funds for data factors/ workflows	Publish Data / sharing data is not currently part of KPIs	Too much emphasis on details instead of foundations	Focus on tools, not on outcomes	Current tools not focused on making life easier	Lack of incentive to standardize and publish data	No dedicated team in each center to standardize data
Some scientists act as if they own the data (but data comes from farmers and is gathered by enumerators); this mindset must change		Improper turnover of data	Some guidelines are optional				

What does a Data Harmonization Guidelines Document need to do?

And what does it not need to do?

Crowdsourced answers during the workshop

- Get buy from Managementt - Program Directors (endorse, enforce)
- What it is, why?
- Successful examples (could be a preamble)
- "You only get $$ if adhering"
- Why this time is different
- Cover letter to directors
- Distinction b/w the Guidelines document, and "lobbying" aspect
- Embed into Monitoring F/W (E.g. in PRMS for accountability)
- Start with W1 / W2, bilaterals later
- Question around how AI fits in: can AI do the harmonization without our own Guidelines?
- Any paper that is published also should have data published
- Clear definition of how CGIAR wants to operationalize FAIR
- Minimum Metadata Values
- Clear "How to" guidelines with tools and vocabulary/ ontology identified
- List of incentives from SO (reward and “punishment")
- Various datasets with common variables ready for re-use, involve key data persons as much
- Don'ts: restrict data accessibility
- Emphasis on open access and transparency

What is needed to be able to start using a Data Harmonization Guidelines Document? What conditions must be met?

Crowdsourced answers during the workshop

Establish KPIs	Identify champions	Create a group of champions, 1 per center, train them	Find funding	Set criteria for "high-value" data sets	Get endorsement from managers	Mapping minimum variables to relevant use cases	Identify 1 tool to use to standardize data
Have a clear vocabulary/ ontology	Organize a Harmonize-a-thon, over a month where the Center with the most datasets standardize wins a prize		Have a clear workflow on how to harmonize data: which tool to use, which vocabulary, support person, where to publish the harmonized dataset, with which supporting documents		Buy-in from Data champions per center	Make the CGIAR Data Harmonization document part of proposal writing process (following the common variables, data management and access plans etc)	

Data Harmonization Workshop
Appendix

Workshop Agenda

Demo Day
Monday / June 16th

Remote Only

5:30 to 8:30 **CGIAR Data collection and cleaning tools**

Day 1
Tuesday / June 17th
Room: **VC-09**

Remote Only

6:00 to 7:00 **Kick-off for Remote Participants** / Remote

Morning / In Person

9:00 to 9:30 **Welcome and Introduction**

9:30 to 10:30 **Demo Day Recap**

10:30 to 11:30 **Agree on Definitions**

11:45 to 1:00 **Review Existing Data Protocols**

Lunch / 1 to 2pm

Afternoon / In Person

2:00 to 3:45 **Minimum Variables** (session 1)

4:00 to 5:00 **Outline of Data Harmonization Guidelines**

Day 2
Wednesday / June 18th
Room: **Sasakawa**

Remote Only

6:00 to 7:00 **Early Morning Summary Call** / Remote

Morning / In Person

9:00 to 9:30 **Intro to Day 2**

9:30 to 10:30 **Ontology for Increasing Fairness and AI Readiness**

10:45 to 1:00 **Minimum Variables** (session 2)

Lunch / 1 to 2pm

Afternoon / In Person

2:00 to 3:30 **Minimum Variables** (session 3)

3:45 to 5:00 **Draft 1 of Data Harmonization Guidelines**

Day 3
Thursday / June 19th
Room: **Sasakawa**

Remote Only

6:00 to 7:00 **Early Morning Summary Call** / Remote

Morning / In Person

9:00 to 9:30 **Intro to Day 3**

9:30 to 1:00 **Best Practices**

- Data Capture
- Data Quality
- Data Cleanup
- Data Integration

Lunch / 1 to 2pm

Afternoon / In Person

2:00 to 2:30 **Reach Final Agreement on Variables**

2:30 to 3:30 **Draft 2 of Data Harmonization Guidelines**

3:45 to 5:00 **Data Harmonization Workplan**

Day 4
Friday / June 20th
Room: **Sasakawa**

Remote Only

6:00 to 7:00 **Early Morning Summary Call** / Remote

Morning / In Person

8:00 to 8:30 **Intro to Day 4**

8:30 to 10:45 **v1 of Data Harmonization Guidelines**

11:00 to 11:30 **Sign off on Data Harmonization Guidelines**

11:30 to 12:00 **Closing Remarks and Reflections**

Lunch / 12 to 1pm

End of Workshop

1:00 to 4:00 **Living Labs Data and Digital Twin**
Side Event
Room: Sasakawa

- Intro & Living Labs Presentations / 60m
- Digital Twin Framework / 30m
- Team Discussion / 60m
- Next Steps / 30m

/ End of document