

Request for Proposal to assist IFPRI in  
AgPile PoC and MVP implementation  
from Jan – Dec 2025

**Questions & Answers**  
(Updated Feb 25, 2025)

| Question   | Answer  |
|--|---|
| Is the deadline for submitting all the relevant material (technical and financial) the 15th of February?   | The proposal submission deadline has been extended to February 28.  |
| Can you forward to us a copy of IFPRI’s Standard Agreement for Services to look into it?   | <a href="#">Research Agreement IFPRI.docx</a>   |
| Do you look for a company that would build 6 POC/MVP or a company that would do one POC/MVP?   | No, we are not looking for one company to build 6 PoCs. Vendor would choose one of the catalogs mentioned in the RFP (most likely the vendor's solution). Vendor would then propose using this solution for the Agpile PoC and MVP implementation. using that as the basi for the AgPile tooling.   |
| Would you buy this service in France or in the US?   | No preference.  |
| Could we involve an IBM Business Partner or do you expect an IBM only answer?  | No preference; answer(s) from Business Partner(s) are okay too.   |
| Will it be a fixed-price contract or a cost-reimbursable contract?   | The contract will be a cost-reimbursable contract to align with IFPRI’s own reporting requirement with the donor.   |
| If I understand correctly, you are looking for a company to support you in the evaluation of 6 POCs, including our solution. So, if we support you, we are judge and jury. Am I right? | No, we are not looking for a company to do the evaluations of the vendor solutions listed. We expect each vendor to propose a solution for AgPile from the candidate solutions listed in the RFP. We will select a vendor from the proposals to collaborate on a proof of concept. Future engagement will be determined after a successful demonstration of the capabilities listed in the RFP. |

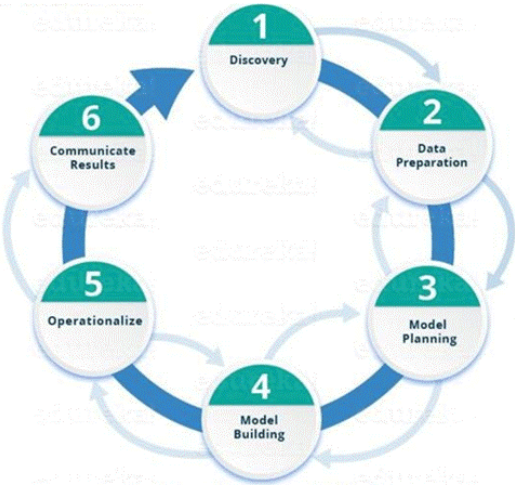
| Question   | Answer  |
|--|---|
| Image assets: is IFPRI providing the assets or do we need our own library for this?  | Image assets will come from the selected research programs involved in AgPile PoC & MVP, not IFPRI.   |
| Where is the tenant to do all what is required? Is that a separate greenfield for AgPile or somewhere in the existing CGIAR landscape?                                     | Please, assume a separate greenfield for AgPile, either within existing CG landscape, or a fresh AgPile environment. Also, if the vendor makes a strong case to host the tenant within the vendor environment during the PoC for flexibility and efficiency, that would be taken into consideration in the evaluation of the technical solution   |
| What is the level of support we will have from IFPRI? Please specify for agritech experience, data platform engineering and CG research expertise (to steer input engine)? | <p>IFPRI (and the leaders within the selected Research Programs) will provide whatever technical, product, domain expertise needed by the vendors to execute the PoC and eventually the MVP per vendor proposal response. The actual number of CG resources/experts from the selected research programs to participate has not been finalized. Recommendations of the level of staffing from CG for efficient execution of PoC and MVP are welcome.</p> <p>The program will build a scalable governance structure for managing data assets and solutions independent of the vendor. The vendor will need to rely on this structure.</p> |
| Where will be the IFPRI team for AgPile based? Are there multiple time zones that we need to cover?  | IFPRI and other CG team members will be based out of their primary work location – Washington DC – USA (IFPRI), Texcoco – Mexico (CIMMYT), Cali – Colombia (CIAT), Montpellier – France (CGIAR) and some selected regional locations in Africa (e.g., Kenya, Ethiopia), East Asia (e.g., Taiwan). Yes, the team will be in multiple   |

| Question  | Answer  |
|---|---|
|   | time zones, but CG will work to minimize the time-zone impact on the vendor.  |
| Does IFPRI have any preference on where the project team is located?  | No preference, since most interaction will be virtual   |
| Data should move as little as possible. But where is the data exactly located? (place and technology, type of database and accessibility) | <p>Yes, the plan is to minimize data movement, both to ensure data owners maintain control of their data assets, and when there is sensitivity with the data. Decision about data movement will be made on a case by cases bases – to address performance issues, when there is a need to protect operational systems, and when data is retrieved from archive or backup, among other reasons.</p> <p>Currently data is located on AWS, GCP, Azure blob stores, SQL and No-SQL databases, on-prem data sources accessible via VPN and FTP, and other technologies. Cloud data stores are in multiple zones all over the world, and on-prem data sources are in data center in countries with limited cloud availability and where there is data localization requirement.</p> |
| Are there existing APIs in place?   | Yes, there are some data assets accessible via API and microservices. More will be needed.  |
| Is there any information around current data maturity in IFPRI?   | CABI did a FAIR assessment of digital assets at CGIAR, and CGIAR has a tool (FAIRscribe) that automates the FAIR assessment. We should assume we will encounter different levels of data maturity, and would need to mature the data assets to make them AI ready.  |

| Question  | Answer   |
|---|--|
|   | <p>Data quality and standardization is quite limited across CGIAR. Therefore it is the core objective and first priority of a scalable governance structure to engage all CGIAR institutions involved in prioritized PoC and MVP solutions to agree on data formats. The project engages experts who have worked with CGIAR scientists on similar challenges to drive this forward.</p>  |
| <p>Are there any Data ops or Dev ops processes in place?</p>  | <p>No formalized Data Ops and Dev Ops. The technical teams supporting the research programs have these processes at different levels of maturity.</p>  |
| <p>The total budget is 1.5 mln, but that is including the mvp phase. How much does IFPRI want to free up per POC? And should the 1.5 mln also cover IFPRI effort?</p> | <p>Suggest using the 1.5MM as a guide to determine the scope of work/ effort to be done by the vendor. We are assuming the PoC as a Mini-MVP to prove critical capabilities before diving into the MVP. PoC should consume no more than 25% of the budget. 1.5MM is the allocation for the vendor effort, it does not include the effort of IFPRI or other participants from CG and other partners. These said, vendors are expected to recommend a budget that the vendor feels would allow them deliver the PoC and MVP, if selected. If possible, provide the case for recommendation of budget increase.</p> |
| <p>Is there a chance multiple vendors + system implementers will be selected for POC? (so e.g. us + vendor and another implementer + vendor)</p>                      | <p>No, unless these vendors /system implementers partner to submit a joint response to the RFP that is considered the strongest and selected to proceed. We are not planning a multi-vendor solution PoC.</p>  |

| Question   | Answer   |
|--|--|
| <p>Are we allowed to ask more clarification questions after the formal deadline?</p>   | <p>Yes, we will continue to respond to questions after the formal dateline.</p> <p>The target is to integrate the vendor as an equal partner into the product management framework as soon as possible to ensure a transparent dialogue and decision-making.</p>   |
| <p>Will all questions and answers from all vendors be shared with all vendors?</p>   | <p>Yes, all questions and answers will be shared with every vendor, but anonymized.</p>  |
| <p>Are there any security considerations we should be aware off during POC?</p>  | <p>No specific security considerations during the PoC</p>  |
| <p>Will the RFP response be evaluated based on POC or MVP approach or both? (i.e. should we focus on explaining the POC in most detail, and MVP on high level because that depends on the POC)</p> | <p>On Both, especially efficiency of scaling/expanding/transitioning the PoC to support agile build out of the MVP</p>   |
| <p>Are there examples of AgPile in existence that could guide the vendors?</p>   | <p>Yes, there are many examples of Apgile in existence. Take a look at these examples.</p> <ul style="list-style-type: none"> <li>• NIH SCHaRE – NIH Science Collaborative for Health Disparities and Artificial Intelligence bias Reduction (SCHaRE)</li> <li>• NiH AllofUs Research Hub – federated digital assets for researchers on working on programs within NIH such as AIM-AHEAD -- Researcher Workbench – All of Us Research Hub.</li> <li>• Open Science Grid – participatory computational environment science researchers (OSG   A national, distributed computing partnership for data-intensive research)</li> <li>• Re3data.org (Home   re3data.org) registry of repositories with research data</li> </ul> |

| Question   | Answer  |
|--|---|
|  | <p>(however has no workspaces and indexing is limited to the repository level, which Agpile will go down to the level of the assets (datasets, pipelines, models, etc)</p> <ul style="list-style-type: none"> <li>• Global Biodiversity Information Facility (GBIF) -- free and open access to biodiversity assets including workspaces, tools, datasets. Tutorials, etc.</li> <li>• VODAN Africa (Virus Outbreak Data Alliance Network in 9 African countries) used COVID-19 research.</li> <li>• Genomic Data Infrastructure (GDI) has a federated catalogue for Bioinformatics Researchers.</li> </ul> |
| <p>Will there be a separate computational environment or is the focus on enabling execution of the workflows created by the research programs on shared computational environment created to support sharing and collaboration with the assets from the research programs?</p> | <p>We expect the solution to include a compute component where workflows, queries and analyses can be executed.</p>   |
| <p>Is there a limit to the partnerships the vendors can get into? Can we partner with more than 1 vendor to submit a response?</p>   | <p>No limit. Yes, you can partner with more than 1 vendor especially if an integrator is combining multiple tools.</p>  |
| <p>Will the tools and assets created by the different research programs be replaced by what is available through AgPile?</p>   | <p>The strategy is to create tools for the selected workstreams (Artemis, ClimMob, Breeding Resources) with the goal of getting additional programs to adopt AgPile as it aligns with them.</p>   |
| <p>Some tools and open source solutions have vendor managed versions e.g., Kubernetes, MLFlow, Delta-Share, Spark, etc. If a vendor is proposing a custom managed version versus a vendor managed</p>  | <p>Yes, any context that a vendor can give context for their recommendations for solutions is helpful because we expect to get platforms that range from fully managed solutions to custom. Considerations will need to be made</p>   |

| Question   | Answer  |
|--|---|
| versions, should we provide our perspective on the preference?   | between portability of solutions versus the cost of operating unmanaged solutions.  |
| Our solution will focus on infrastructure agnostism so to ensure ability to broad interoperability. If reuse is a major consideration, this would allow existing processes on any cloud or on prem to be integrated into AgPile with minimal modification. How important is the ability to take advantage of existing assets e.g., Artemis pipeline on GCP, Bioflow workflow on Azure, Agronomy models on AWS, etc with minimal modification or replatforming? | Potential re-use of existing assets and tools will be considered especially as they may relate to ease of adoption for our initial workstreams. Please note that reusability of existing assets and tools will be one consideration among many that will inform vendor partner selection.   |
| Could you confirm the scope of the agreement signed, please? Is it only on the POC? Do we need to provide a technical and financial response for the POC part only or also for the MVP part? Will we sign a new agreement for the MVP?   | There will be one agreement that covers both POC and MVP, but the POC must meet our expectations to proceed to the MVP.   |
| <p>Regarding the perimeter of the POC/MVP, could you please confirm the following points:</p>  <p>Which phases among those in the diagram below should be addressed in the PoC and/or in the MVP?</p>   | <p>Each phase in the diagram should be addressed (i.e. discovery to ingestion of raw data to data curation to creation and hosting of analytical outputs, models, data products). However, the individual workstreams will be responsible for developing their processing pipelines. To put it another way, the vendor will help the AgPile team deliver a platform that the workstream projects can use in order to operate each phase of the data cycle described in the diagram.</p> |



| Question  | Answer  |
|---|---|
| <p>What are your deployment expectations (do all the countries listed in the RFP need to be "plugged in" in the PoC?)</p>   | <p>A single region deployment would be fine for POC. Beyond the PoC, multi-region deployment will need to be considered based on usage and performance.</p>   |
| <p>In the POC, are you considering financing the on-cloud platform + services?</p>  | <p>Yes, we can consider financing the on-cloud platform and services for the POC purpose.</p>   |
| <p>Can you please confirm or correct our understanding and main assumptions below, for the PoC and the MVP ?</p> <p><u>Data storage &amp; access :</u></p> <ul style="list-style-type: none"> <li>• The expected platform will store images &amp; metadata, when allowed.</li> <li>• The expected platform will federate remote sources (eg : images &amp; metadata restricted to a specific country), when required.</li> <li>• The expected platform must provide a single entry-point, for all users WW, with restrictions on available data.</li> </ul> <p><u>Data sovereignty :</u></p> <ul style="list-style-type: none"> <li>• Some data (images &amp; metadata) cannot be stored nor consumed/viewed outside of their original country.</li> <li>• The 'countries with restricted data' need to do cross-analysis based on their own restricted data + data from other 'no-restriction countries'.</li> </ul> <p><u>Data preparation :</u></p> <ul style="list-style-type: none"> <li>• The tagging &amp; classification of existing images will involve CV capabilities, images will be tagged once uploaded to the dedicated platform.</li> </ul> | <p><u>Data storage &amp; access :</u></p> <ul style="list-style-type: none"> <li>• The expected platform will access and index images and metadata stored in the participating content repositories for research programs, research centers, or other relevant public sources.</li> <li>• The expected platform will enable federated access to participating content sources, so to minimize the need to consolidate content from multiple sources into a single location, allowing access to content where they exist to the extent allowed during processing to achieve acceptable performance.</li> <li>• The expected platform will allow unified user access interface for all users worldwide, restrictions on available data to be applied as role-based access control</li> </ul> <p><u>Data sovereignty :</u></p> <ul style="list-style-type: none"> <li>• Yes, let's assume there will be localization restriction on where some data is stored and used as more sensitive information is indexed and made accessible to users. For the PoC and MVP, localization restriction, if any, will</li> </ul> |

| Question  | Answer  |
|---|---|
| <ul style="list-style-type: none"> <li>The expected platform has to offer technical capabilities for data cleaning &amp; refining.</li> </ul> <p><u>ML &amp; Datascience :</u></p> <ul style="list-style-type: none"> <li>The expected platform has to offer technical capabilities for Machine Learning (classic &amp; GenAI) &amp; DataScience.</li> </ul> <p><u>Data visualization :</u></p> <ul style="list-style-type: none"> <li>We will implement a custom User Interface with a search feature to browse images.</li> <li>Other advanced visualization (maps, charts, ..) will be done in notebooks.</li> </ul> | <p>mostly relate to storage of the data, not consumption or viewing.</p> <ul style="list-style-type: none"> <li>Country level analysis in countries with data restriction will be based on its own data. If there is a need for comparative analysis and benchmarking with other countries, only unrestricted data from the other countries will be used to ensure compliance with data use agreements and licenses.</li> </ul> <p><u>Data preparation :</u></p> <ul style="list-style-type: none"> <li>Tagging and classification of images is part of CV effort within the research programs. However, tagging and tag updates will happen as many time as needed by the various CV/digital phenotyping use cases applicable to the images</li> <li>The expected platform has to provide tooling for additional data handling activities as well as shared workspace needed to process the data for AI readiness.</li> </ul> <p><u>ML &amp; Datascience :</u></p> <ul style="list-style-type: none"> <li>The technical capabilities for Machine Learning (classic &amp; GenAI) &amp; Data Science capabilities are needed. Also, an AI assistant to accelerate FAIRness.</li> </ul> <p><u>Data visualization :</u></p> <ul style="list-style-type: none"> <li>UI with advanced search feature to FAIRify relevant content (including images) in the repositories based on researcher interest. Also, an AI Assistant to accelerate FAIRness.</li> </ul> |

| Question   | Answer  |
|--|---|
|  | <ul style="list-style-type: none"> <li>Advanced visualizations can be done in notebooks, as well as in other visualization libraries and tools, per researcher’s preference.</li> </ul>   |
| <p>Regarding the requirements “automated annotation and labelling of images at the point of capture” and “Harvesting, curation, and labeling of images for CV model development,”</p> <ul style="list-style-type: none"> <li>Do you already have an AI Model trained to analyze vegetable or fruit images and able to automatically provide information (e.g., species, maturity level, diseases, etc.)?</li> <li>Do you ask to implement in the PoC or in the MVP an end-to-end solution (e.g., with a Smartphone application, to capture the images in the fields, that will be processed by an AI model, in order to classify and tag the images)?</li> <li>Regarding “Demonstration of federated access and distributed execution of processes as needed.”, can you please explain in detail the kind of processes you are expecting?</li> </ul> | <p>Yes, there are AI models in place and in development for computer vision digital phenotyping. Also, the use of Roboflow to automate annotation/labeling AI model training is in progress. Currently, the execution of these AI models requires the images to be consolidated into a single storage location for execution. The PoC will involve enabling the execution of these annotations and labeling AI models against image stores in a federated manner, without the need to consolidate the image into a single store. The PoC demonstration of federation will be the generation of the image annotation file within the image store or at a nearby workspace associated with the image stores located where the annotation process is unable to generate associated annotation files, and indexing these image stores and associated annotation files for Findability, Accessibility, Interoperability, and Reuse by CV digital phenotyping researchers.</p> <p>PoC &amp; MVP end-to-end solution would focus on enabling FAIR on these assets such as apps, calculators, digital tools, workflows, models, datasets, etc. The actual creation/build of these assets such as creating smartphone apps, capturing images in the field, and building the CV models for image processing will all continue to be the work of the researcher/ research groups. An important deliverable of the PoC is the guideline for the</p> |

| Question  | Answer   |
|---|--|
|   | <p>Researchers to follow to create digital assets to ensure FAIRness within AgPile. This deliverable will be the connection between digital assets created by the Research Teams and AgPile.</p> <p>For demonstration of federated access would involve showcasing the ability to access participating assets across domains, repositories, etc. with a unified set of credentials, and federated / distributed processing would involve showcasing the ability to federate/distribute processing across multiple workspaces associated with the participating assets to limit the movement of data to results required by the process. These are critical capabilities for AgPile to take advantage of the infrastructure resources of the participating research programs/research centers. Also, the frameworks adopted to enable these capabilities are up to the vendor and solution to make these capabilities seamless for users and collaborators in AgPile.</p> |
| <p>Can you confirm that there is no central identity management solution? Are we correct that each research center may have their own systems (e.g Active Directory, Okta, etc.)?</p> | <p>We anticipate that identity management will be federated, with each research center or research program providing identity provider that is unified in AgPile for SSO and federated access. We anticipate that not all Research Centers and Research Programs currently operate an identity provider. Those Research Centers and Research programs will be supported to set one up so it be integrated into AgPile identity management capability.</p>  |
| <p>Is there an existing data standard for anonymized household data collection currently in use? How is</p>   | <p>There are various approaches in use by Research Centers and Research Programs</p>   |

| Question   | Answer   |
|--|--|
| <p>this information captured (ie from the users of the mobile app taking images of plants)?</p>  | <p>to anonymize sensitive data, including masking, aggregation, and generalization, and less so perturbation, data swapping, data synthesis, pseudo-anonymization and differential privacy.</p> <p>Sensitive data is also captured in various ways – through surveys (manually and electronically), through research tools and applications (both mobile and web-based)</p>  |
| <p>What is meant by labeling and annotating? In terms of existing image data, is some labeled and annotated and some not? Is manual metadata tagging required?</p> | <p>By labelling and annotation we mean adding metadata and descriptive information to multi-media content (images, videos, audios) to make them easier to understand, organize, and analyze by humans and machines.</p> <p>Specifically,</p> <ul style="list-style-type: none"> <li>• Labelling involves (1) assigning pre-defined categories and classes to different parts of the multi-media content, (2) adding tags and keywords to describe their main features or topics, and (3) establishing boundaries around the object of interest to establish their location and size in the object.</li> <li>• Annotating could be (1) text annotation – adding text (2) audio transcription – transcribing audio to text (3) segmentation – breaking down content into parts / regions, (4) attribute annotation – adding specific info like color, shape, behavior, etc., (5) KeyPoint annotation – marking specific points on the image or video</li> </ul> <p>Some existing images are labelled and annotated, but a lot more are yet to be</p> |

| Question   | Answer   |
|--|--|
|  | <p>labelled and annotated. Lots of the labelling and annotating has been manual. Only recently have automated labeling and annotating tool like Roboflow been introduced. We want to scale automated labelling and annotation, as well as training labelling and annotating models for new images, at the point of image capture and storage.</p> <p>Manual metadata tagging may be needed by Breeding and Agronomy Scientists. The opportunity for AgPile it is to scale and accelerate labelling and annotation with AI.</p> |
| <p>Which research centers are currently using ML and are they sharing models with other centers?</p>   | <p>All research centers are using ML and are sharing the models ad-hoc based on personal relationships and contacts. AgPile will to provide an environment for the sharing and collaboration with these assets, and the "connective tissue" to unify FAIR* Federated AI Readiness. (*Findable, Accessible, Interoperable, Reusable)</p>  |
| <p>Do you provide already anonymized data, or do you need the PoC/MVP platform to provide a feature to anonymize the datasets?</p>   | <p>Both options should be assumed. For assets resources where data has been anonymized already the PoC/MVP will take advantage of that. For those that still require anonymization, platform should consider a feature to support that requirement.</p>  |
| <p>Have you already chosen to use Roblofow for the image labelling tool &amp; AI model for the PoC &amp; MVP? if yes, will you have a full Roboflow environment active and usable for the PoC/MVP?</p> | <p>Roboflow is in active use as one of the tools for CV model work. Let's assume the Research programs (EIB (Artemis, 1000Farms) , Tumaini) that use Roboflow will have a shared environment that we can leverage for the PoC/MVP as needed.</p>   |

| Question   | Answer   |
|--|--|
| <p>Regarding “federated / distributed processing”, can you please give us some examples of these processes (e.g., Metadata retrieval? Creation of an AI model on top of federated datasets? What other processes?)</p> | <p>Some Examples:</p> <ol style="list-style-type: none"> <li>1) federated/distributed processing – a) listing germplasm records (including images) by breeding trials sites hosted in multiple gene-bank repositories, b) federated dataset of agritrials by variety with agronomy, breeding, environmental, soil, climatic, etc. Attributes.</li> <li>2) metadata retrieval – consider this as the process of collecting or harvesting data about the data (and all other assets) from the participating repository resources that is indexed by AgPile to unify FAIRness. We will also encourage metadata publishing for participating repositories to actively manage the state of their assets on AgPile.</li> <li>3) creation of AI model on top of federated datasets – CGIAR has lots of Agric AI models and underlying datasets that are in different locations, AgPile intends to index these models and datasets, so the models can be extended and datasets can be put to secondary use to derive new insight. Examples are Computer Vision models for Digital Phenotyping, Labelling and Annotation Models etc.</li> <li>4) Other processes – AgPile would index all pipelines for data processing, ML and AI development, released by the Research Programs and Centers for Community Use. Let's think of indexing all relevant digital assets, instead of just processes or datasets.</li> </ol> |
| <p>How many distinct users will need to connect onto the platform during PoC and MVP?</p>  | <p>For Year 1 we are targeting Researchers focused on digital phenotyping acceleration of breeding, agronomy,</p>  |

| Question   | Answer  |
|--|---|
|  | <p>climate change, gender studies and trials. If we can mobilize up to 50 researchers from the key breeding and agronomy agtrial sites and CGIAR priority countries to actively use the platform that would form a great foundation to grow from. Long term, AgPile should have hundreds of thousands of users and collaboratives from around the world.</p>  |
| <p>How many image files and how many data files will have to be considered, during the PoC and during the MVP?</p>   | <p>We do not know exactly how many image files and the size of the image store, but we can assume it would be big-data. TumainiApp has a dataset of 55K banana images and would like to expand the banana image storage to 5 TB and more. Artemis has reported about 40TB of image storage and it's growing rapidly. Please assume that all these assets will be the different research program and research centers storage location and indexed for FAIRness in AgPile.</p> |
| <p>Is the goal of the catalog to facilitate access to the data stores to all interested parties within and external to CGIAR or to provide a data governance structure to manage the access to the data assets?</p>  | <p>Yes, the goal of the catalog is to enable FAIR (findability, accessibility, interoperability and reusability) of assets to both internal and external users. The DG structure would be embedded in the assets to ensure FAIR.</p>  |
| <p>The RFP discusses parties both internal and external to CGIAR to be able to find and leverage assets. Is it expected that external users would use the data assets through their own analytic tools or will all users (internal and external) be expected to use the CGIAR platform to analyze and use the data assets?</p> | <p>Both options. External users should be able to request to connect their workspace(s) and asset(s) to AgPile for sharing and collaboration. Also, external users should have flexibility to use CG workspaces and tools as well as their own analytic tools, with the understanding of the requirement to share the output of their analysis either</p>   |



| Question   | Answer  |
|--|---|
|  | in a publication or an app accessible on the web / mobile.  |
| Do the current solution uses cases apply to both internal and external or just internal CGI users?   | Current use cases apply to both internal and external "potential" users. However, initial focus of the PoC and MVP is internal CG users   |
| Do all internal CGIAR users have Azure Entra IDs to access a common analytics platform? Is it ok to have this or is this a requirement for all parties to have IDs for access? | No. It is safe to assume they all do not have IDs, but if the federated identity provider solution requires this ID this should be called out in the proposal response.                                     |
| Which organization will sign the agreement – IFPRI or CGIAR System Organization?   | The agreement will be signed by IFPRI on behalf of CGIAR. While IFPRI is a CGIAR Research Center, IFPRI (based in the U.S.) and the CGIAR System Organization (based in France) are two different entities. |