Advisory Services

CGIAR

CGIAR Research Program 2020 Reviews: WHEAT - List of Annexes

CGIAR Research Program 2020 Reviews: WHEAT - List of Annexes

Annex 1: Terms of Reference for the CRP 2020 Review

TERMS OF REFERENCE & CALL FOR EXPRESSIONS OF INTEREST

CRP 2020 Independent Reviews of Quality of Science and Effectiveness

Background

In 2020, the CGIAR Advisory Services Shared Secretariat (CAS Secretariat[1]), through its evaluation function, is planning independent reviews of the twelve CGIAR research programs (CRPs[2]). The reviews, commissioned by the CGIAR System, will provide information on Quality of Science and Effectiveness in each CRP. The CAS Secretariat has been mandated to undertake this work as part of its role in providing independent evaluation and assessments to the CGIAR System[3]. The reviews are designed to be rapid (completed within 11 weeks) and produce top-level findings, but not to generate the range of in-depth information as would be obtained from an evaluation. Further, the reviews are entirely desk-based, and no travel is planned.

Between April and December 2020, teams of two external expert consultants will each review one CRP, relying on its documentation and a limited number of virtual (telephone or online) interviews with the CRP Program Leader, staff and key external stakeholders. An internet-based survey will also be conducted for CGIAR researchers and CRP donors and partners. Bibliometric analysis conducted by the CAS Secretariat will supplement the information available to the expert reviewers.

The CRPs were designed to run for six years, from 2017 to 2022, but have been curtailed by one year and are now scheduled to conclude in 2021. Each CRP is composed of 3 to 5 Flagship Programs (see Annex 1), which in turn operate clusters of activities for research. The CRP reviews will rely on data and information available for the period 2017-2019, and will inform future research modalities to be developed in 2021.

A key document for the CRP review is the program Theory of Change, which in many cases may be the version developed in the CRP proposal or its updates. In some programs the Theory of Change may be implicit or not completely documented. The external experts who will conduct the reviews will rely on additional sources (annual planning documents or interviews) to understand the Theory of Change in use by the CRP, which will be the basis against which the program will be reviewed. The Flagship Programs within the CRP each have their own Theories of Change, which are nested under the CRP Theory of Change. Together, the hierarchy of the CRP and Flagship Theories of Change form the key reference documents for the CRP 2020 Review.

As a desk-based review, this effort will attempt to minimize the burden on CRPs. In advance of the reviews, CRPs will prepare the set of reference documents for the review. At the start of each review, the CAS Secretariat will organize an initial briefing involving the team of expert reviewers and the respective CRP Lead and staff .During the data collection phase, the review team will conduct an interview with the CRP Leader and a focus group discussion (FGD) with other members of the CRP management. The review team will provide a debrief discussing the preliminary findings with the CRP management and the CAS Secretariat, for validation and feedback. The draft report will be shared with the CRP Leader and staff for factual correction and final feedback. CRPs may choose to provide a formal management response to the review, though this is not a requirement.

[1] See Annex 4 for a list of acronyms used in this Terms of Reference
[2] See Annex 1 for a list of the twelve CRPs and their associated Flagship Programs.
[3] The CAS Secretariat/Evaluation 2021 workplan will propose a similar review or evaluation of the CGIAR Platforms, creating a harmonized Terms of Reference that has been adjusted to Platform's characteristics and function. The Platforms are considered separately from CRPs in order to address aspects of their work that differ substantially from the CRPs.

In July 2020, the CAS Secretariat also will conduct an after-action review with the Program Leader and staff from the first three CRPs reviewed, to ensure that the approaches used to pursue the review questions are as streamlined and appropriate as feasible.

Purpose of the review

The primary purpose of the CRP 2020 review is to assess the extent to which CGIAR research programs are delivering Quality of Science and demonstrating effectiveness in relation to their own Theories of Change (or other planning documents stemming from the Theory of Change set forth at program inception, in the event that the original Theory of Change has not been updated to reflect the current thinking behind the CRP's work). Within that primary purpose, the objectives of the independent CRP reviews are captioned below:

1. To fulfil CGIAR's obligations around accountability regarding the use of public funds and donor support for international agricultural research;
2. To assess the effectiveness and evolution of research programs' work under CRP 2017-2021;
3. To provide an opportunity for programs under review to generate insights about their research contexts and programs of work, including lessons for future CGIAR research modalities.

Expected uses and users of the CRP 2020 reviews

The CRP 2020 reviews are a key step in the CGIAR System's demonstration of accountability. Accordingly, the primary users of the reviews will be the CGIAR System Council, with insights and lessons developed from the reviews for use by the programs themselves.

Recognizing the potential of these reviews to support Program Leaders and their teams, the CAS Secretariat will engage the expert review team to work with each Program Leader in defining any supplementary questions of specific interest to their CRP, which will be included in the scope of work for the respective CRP review, subject to the limitations of time and resources for the review. Interested consultants should keep in mind that the final scope of work follows the structure and process laid out in this Terms of Reference and for some CRPs may include 1-2 well-defined additional question(s) from the CRP under review.

Further, the CRP reviews may provide lessons that inform the transition to One CGIAR in 2022, based on the program-level findings and a synthesis of system-level findings in 2021; to that extent, the reviews will be a future reference for system management in the change process.

In the final report, the expert review teams are expected to identify findings, conclusions and recommendations that apply to CRPs for use in refining the 2021 Plans of Work and Budget (POWB) to the extent feasible in the remaining program year, and lessons to inform future research modalities.

Scope of the CRP 2020 review

The CRP reviews will cover 12 CGIAR research programs from the proposal acceptance date in 2017 through 2019, making use of all the reporting and monitoring information available to date. The first three reviews will rely on the CRP's 2019 draft annual reports, prior to their vetting and quality assurance by the CGIAR System Management Office, and the other nine reviews will use the finalized CRP annual reports. The scope will include the program of work of each CRP and its Flagship Programs, with the reviews guided by the CGIAR's Quality of Science and Effectiveness criteria, and the Theories of Change for the CRP and its Flagship Programs. The reviews will not assess individuals, teams, or institutes in which programs reside. Emphasis will be on the CRP's Sphere of Control, that is, the quality of inputs, activities and outputs, and Influence, that is, short and intermediate outcomes that are expected to lead to a development impact.

The CGIAR System defines outcome-level changes as Intermediate Development Outcomes (IDO) and System Level Outcomes (SLO), as described in detail on the CGIAR website[4]. The CRP 2020 Reviews will focus on the IDOs, including sub-IDOs, given the short span of time (three years) for the current phase

4 https://cas.cgiar.org/sites/default/files/ISPC_WhitePaper_SLOsIPs.pdf

of CRPs. Expectations of documented outcomes will be informed by (a) the amount of time the research has been conducted under the CGIAR and its centers, including research prior to the CRP in the case of legacy programs, and (b) whether the CRP's targeted first users of research outputs are within the research community or closer to market adoption. It is not expected that all planned outcomes will have been achieved by the CRPs at the time of its review, because the present reviews are to be conducted after three years of operation on five-year research programs (originally planned for six years). Where data on impacts have been reported in an Outcome and Impact Case study Report (OICR) these will be included in the review. To the extent feasible, the review of CRP effectiveness should assess the likelihood for achieving IDOs and/or sub-IDOs, based on the CRP's and its Flagship Program's documented performance in relation to their Theories of Change.

Review Criteria

The CRP 2020 Review will be based on two of the six CGIAR evaluation criteria as defined in the CGIAR Evaluation Policy[5], which comprise relevance, quality of science, efficiency, effectiveness, impact, and sustainability. Because the CAS Secretariat/Evaluation Function has been directed to execute the external reviews in a compressed timeframe, the two criteria for assessing the CRPs that have been agreed with the System Council committee that is concerned with evaluation are Quality of Science and Effectiveness.

Quality of Science in the CGIAR is defined as the ways by which research is designed, conducted, documented and managed, in terms of the processes, inputs and outputs. The CGIAR's definition of Effectiveness aligns with that of OECD-DAC and other international bodies as the extent to which objectives have been achieved. An element of effectiveness present in the definition of impact is "a chain of events to which research outputs and related activities have contributed that are likely to contribute to impacts."[6] The application of these criteria in the CRP 2020 Review is further elaborated, below.

Review of Quality of Science

The CRP 2020 Review will examine quality of science and looks both at the conditions that are in place for assuring high quality of science, and the conduct and outputs of research. A systematic and consistent review of science quality across research programs and program components has three dimensions:

- Processes for assuring and enhancing science quality (staff recruitment, performance management and incentives; review processes used; codes of conduct; monitoring, evaluation and oversight for enhancing science quality);
- Inputs (quality of staff and research leaders, facilities and equipment, data management, research design);
- Outputs (volume and quality of publications, genetic materials, etc.).

The above dimensions are captured and elaborated in the review questions, below.

Review of Effectiveness

Assessing effectiveness of a CRP includes documenting the achievement of outputs and outcomes based on program reports and interviews and surveys of people involved or in a position to observe these. Outcomes or impacts will be included when those have been reported in an OICR. The CGIAR reporting definitions of these terms, and a modification made in the definition of outcome for these reviews, are as follows:

- *Outputs:* Knowledge, technical or institutional advancement produced by CGIAR research, engagement and/or capacity development activities. Examples of outputs include new research methods, policy analyses, gene maps, new crop varieties and breeds, institutional innovations or other products of research work.
- *Outcome:* A change in knowledge, skills, attitudes and/or relationships, manifested as a change in behavior, to which research outputs and related activities have contributed.

5 https://cgspace.cgiar.org/handle/10947/2762
6 https://marlo.cgiar.org/glossary.do

For the CRP 2020 reviews, the definition of outcome will be expanded to include innovations[7] that have entered into use. CGIAR defines innovation as follows: "development innovations are new or significantly improved (adaptive) outputs or groups of outputs - including management practices, knowledge or technologies. This could also refer to a significant research finding, method or tool. A significant improvement is one that allows the management practice, knowledge or technology to serve a new purpose or a new class of users to employ it … ."[8]

- *Impact:* A change in state resulting from a chain of events to which research outputs and related activities have contributed. Some examples: crop yield, farm productivity, household wealth (state) income (flow), quality of water (state), water flow (flow).

The CRP 2020 Reviews will assess CRP effectiveness from two perspectives. The first will compare planned versus completed outputs and outcomes as provided by the programs in annual Plans of Work and Budget and Annual Reports for 2017, 2018 and 2019. The second perspective is to assess reported achievements against the CRP's and its Flagship Programs' Theories of Change, which articulates the pathways from outputs to a sequence of outcomes and impact, to be tested in the course of program implementation. As noted earlier, the CRP's Theory of Change is either the original version from its proposal with any updated documentation or, if that Theory of Change has not been followed, an implicit theory in the CRPs annual work plans (POWB). The Flagship Programs' theories of change supplement the CRP Theory of Change as additional reference documentation.

The likelihood of future progress is a further aspect of effectiveness to be examined in the 2020 reviews**.** Whether or not there is a pipeline of innovations, which are reported by stage such as "ready for take up" and policies influenced by sphere of influence, will be determined. Reports of capacities developed, environment enabled, and key partnerships in place for development will also be considered as will opinions of research managers and key partners. Another important factor in future effectiveness, and a common question asked in CGIAR external evaluations, is about the management and governance that is in place in the CRP. Evidence gathered will include presence of a learning environment, addressed and unaddressed challenges to success, and integration across other CRPs' work.

Questions for the CRP 2020 Review

Cross Referencing to the CGIAR Quality of Research for Development Frame of Reference

The CRP 2020 review will cross-reference and map Effectiveness and Quality of Science to the CGIAR's broader Quality of Research for Development (QoR4D) frame of reference. The QoR4D frame of reference encompasses all review criteria and indicators, albeit organized in a different fashion and with a stronger emphasis on how each CRP positions its research and outputs for development outcomes and impact. In deploying two out of six of the evaluation criteria (i.e., as defined in CGIAR's 2012 Evaluation Policy and its accompanying Guidelines), while also bridging with the QoR4D frame of reference adopted by CGIAR in 2017, the review will overtly map the query areas and indicators to the QoR4D frame of reference so that the CRP 2020 Reviews speak to the QoR4D frame of reference. For more information, refer to the QoR4D brief on the CGIAR website: https://cas.cgiar.org/isdc/publications/quality-research-development-cgiar-context

To guide the planning and implementation by the expert review teams contracted to complete the CRP 2020 Reviews, questions for the review have been provided below. These questions were developed based on the definitions of the two review criteria (Quality of Science and Effectiveness), existing self-reported program data and internally funded studies by external experts. This set of review questions will be applied in each CRP review. As noted earlier, the CAS Secretariat will arrange for an initial briefing between the expert review team and the CRP under review, which will include a discussion to define 1-2 supplementary questions of interest to the CRP itself, if any.

[7] CGIAR glossary (https://marlo.cgiar.org/glossary.do) defines an innovation as an output while most research evaluation defines an innovation as a new or improved technology, product, process, or business model that has been put into use (OECD/Eurostat 2005).
[8] https://marlo.cgiar.org/glossary.do

Quality of Science

1. To what extent does the CRP deliver Quality of Science, based on its work from 2017 through 2019?

1.1. To what extent does the CRP benefit from sufficient high-quality inputs (with reference to the research environment and project designs)?

The review should look at productivity and engagement of scientists; diversity of teams and partnerships, in relation to planned outcomes; quality of facilities, equipment and other tools for research; and the level and predictability of CRP funding during the review period.

1.2. To what extent do the CRP management processes ensure the quality of science, including credibility, legitimacy, relevance to next stage users, and potential effectiveness, of the research and operations?

The review will consider the CRP's periodic re-assessment of the demand and quality of research; the research work environment as enabling QoR4D; research ethics, transparency and procedures for conflict of interest; and use of learning mechanisms to inform current and future research, for ultimate users of the research.

1.3. In what ways are the research outputs, such as germplasm, knowledge tools and publications, of high quality?

The review will assess external recognition of CRP outputs as high quality; collaboration for innovation with next stage users and/or beneficiaries; value of outputs in developing capacities for researchers, next stage users and partners.

Effectiveness

2. What outputs and outcomes have been achieved and what is the importance of those identified results?

2.1. To what extent have planned outputs and outcomes been achieved by 2019?

The review should examine the CRP's own targets and deliverables (outputs, milestones, and outcomes) as listed in the program's Plan of Work and Budget (POWB) and annual reports or in the OICRs; as well as contributions to cross-cutting issues, and integrated work with other CRPs.

2.2. What is the importance of achieved outcomes, with reference to CGIAR intermediate development outcomes (IDOs) and sub-IDOs, cross-cutting issues (Capacity Development, Climate Change, Gender and Youth), and partners' objectives, with consideration for predictability of funding and legacy time frame for the CRP?

The review will focus on IDOs and sub-IDOs and other unanticipated outcomes reported by the CRP, whether positive or negative; the program's engagement with cross-cutting issues, namely gender, capacity development, innovation and partnerships; the program's age and maturation (with research in some cases preceding the current CRP cycle) and the context of its work; and achievements in relation to partners' expressed needs.

2.3. How have the program's management and governance supported the CRP's effectiveness in research?

The review will consider changes and adaptations in the program's activities, objectives, and strategy based on lessons learned; unaddressed changes in context or other challenges; and risk management planning and mitigations by the CRP.

2.4. To what extent has the CRP and its Flagship Programs made progress along their Theories of Change?

The review will assess how the program has used its TOC, if at all, or developed an alternative program logic; progressed along the defined impact pathways; and adapted its TOC (explicit or implicit) based on learning and evidence.

Future orientation

3. To what extent is the CRP positioned to be effective in the future, seen from the perspectives of scientists and of the end users of agricultural research (such as policy-makers, practitioners or market actors)?

3.1. What programmatic evidence exists for future effectiveness within the life of the program (through 2021), considering the comparative advantages of the CRP and its Flagship Programs and drawing on the CRP's and its Flagship Programs' progression according to their Theories of Change?

The review will assess the readiness for adoption of the program's deliverables at the IDO and sub-IDO levels; and changes in the program's enabling environment, capacities and partnerships that prepare its research outputs for successful use by next users and beneficiaries.

As noted, 1-2 supplemental questions may be developed by the expert review team and senior scientists and leadership from the CRP under review, with guidance from the CAS Secretariat. These limited questions will align within the three primary review questions as shown above, and will not constitute additional, stand-alone review questions.

Methods and data sources

The reviews will rely extensively on CRP documentation and interviews with Program Leaders and external groups including research partners, national policy-makers and donors, and FGDs with CRP staff. Additionally, bibliometric analysis of CRP research products (publications) will be conducted by the CAS Secretariat and provided to the expert review team. The primary sources of data and information for the reviews comprise the following:

*Documents from the CRP***:** These include CRP proposals (2016-2018) including the CRP's Theory of Change as well as any documented updates or revisions, the CRP's Flagship Program Theories of Change, program independent steering committee reviews, CRP Plans of Work and Budget (POWB), Annual Reports for 2017 through 2019 (the 2019 annual report will be drafted by April and finalized by July 2020), the internal program MARLO data system or the Measurement, Evaluation and Learning Platform of the CGIAR MEL organization, the most recent CRP independent, external evaluation report (for CRPs that had such an evaluation), impact studies from the past five years (for CRPs that have had such a study) and other relevant program documents.

CGIAR Results Dashboard: The results dashboard is an online portal that summarizes each CRP's reported results, including innovations, capacity development, policies and partnerships, as a quantitative supplement to the CRP annual report.

CGIAR database of Output Impact Case Reports (OICRs): OICRs are short reports describing and documenting the contribution of CGIAR research to development outcomes and impact, searchable by geographic location, level of maturity along the impact pathway, or by their contribution to CGIAR's IDOs. The benefit of the OICR analyses is its critical review of the development effectiveness of the CRP's work, in generating lessons learned based on concrete cases, for the design of future research arrangements.

Interviews with CRP Leaders, donors and partners, including CRP Program Directors and levels of management above them. Their wide perspective will be particularly helpful for key accomplishments now and projected for the future and challenges faced. The expert review team will conduct these short (approximately one hour) interviews by phone or video conference call.

Focus group discussion (FGD) with CRP management, to assess aspects of quality of science and the research environment, and to obtain broader views on management and governance. The expert review team will conduct FGDs through a virtual (webinar with video) setting.

External Expert Studies: Any outcome and impact assessment studies conducted or commissioned by the CRP itself, as well as external assessments on other subjects including those that cross-cut programs, are also available.

The CAS Secretariat will conduct pre-analysis on the datasets captioned below, and provide the outputs to the review team for inclusion in the analysis of the CRP. The review team does not need to access these data sources directly. These include data and information from the sources below.

Bibliometric and Altmetric and Other Studies of CRP Publications and Other Outputs: These studies are done mostly by CRP or CGIAR staff. Sources of information about outputs such as datasets, innovations, contributions to policy-making and decision support tools include literature and website reviews.

Survey of Researchers in CGIAR and research partners. To avoid researchers receiving multiple surveys, a master list will be compiled of researchers and the programs/flagships each is involved with. Individual

programs could add a few program-specific questions to the general battery of general interest questions such as opinions of the research environment.

Survey of Partners, defined as a relationship with CGIAR with specific objectives (fund, joint planning or implementation). To avoid partners receiving multiple surveys, a master list will be compiled of partners and the programs/flagships each is involved with. Individual programs could add a few program-specific questions to the battery of general interest questions such as satisfaction with joint efforts with CGIAR.

Overview of Methods and Analysis

These reviews will use mixed methods, quantitative and qualitative, so that analysts can triangulate perspectives, both internal (CRP) and external (partners, next users, etc.) in analysis. When assessing a CRP's quality of science, the expert review team will derive findings from existing CRP documents, bibliometric analysis and reports of any external expert reviews, and from primary data collection from questions on surveys of researchers and partners, interviews with CRP leader (also Principal Investigator) and external stakeholders, and focus group discussions with others in the CRP management.

Publication data collection, bibliometric and Altmetric analyses, and a set of analyses of CRP results are done internally by the CAS Secretariat and CRPs. Analysis of the quantity and quality of research outputs, the number of publications in peer-reviewed journals and other outlets, and the citation of those publications by other scientists will be provided to the expert review team for triangulation of findings.

Three general methods will be utilized in assessing programs on both effectiveness and quality of science: content analysis, descriptive and statistical analysis, and synthesis of existing external evaluations. More information on each of these follows.

- Content Analysis. Quantitative and narrative descriptions of achievements and programmatic actions are found in the CRP documents for the review, particularly the proposal, annual plans, annual reports and selected OICRs. Content analysis of individual reports and cross-report analysis can summarize findings for many of the review questions, including production and utilization of non-publication outputs such as datasets and training events. This could include preliminary analysis of trends observed, given the low number of available data points in the period under review.
- Surveys and Interviews with Statistical and Content Analysis. Representative samples of both researchers and partners will be developed for surveys. Interviews will be done with the CRP manager and selected partners, and an FGD will be conducted with the CRP management and staff. Qualitative analysis will be done on open-ended questions. As with any survey, statistical analysis will be completed with survey responses where that is feasible.
- Synthesis. The content of existing external studies will be aligned with stated objectives of the program and findings in these studies summarized. In a few cases, the studies themselves provide a synthesis across studies to draw more general conclusions.

Methods for documenting the CRP's effectiveness and responses to challenges rely on examining the Theories of Change or alternative program logic at the program and flagship levels in relation to the CRP's reported results from monitoring data (reported on CGIAR's MARLO and/or MEL platforms) and outcome/impact case reports (i.e., OICRs).

- *Comparison of achieved results versus proposed objectives/milestones*. Because each CRP uses annual work plans (POWB) and produces annual reports, it will be relatively straightforward to compare planned outputs against reported completed deliverables (some CRPs may also use milestones, along with or instead of deliverables). The reports also record when deliverable deadlines slip, with explanations for that lack of expected progress. Tagging innovations by stages will also help with year to year comparisons.
- *Comparison of operational or proposed theories of change with reported achievements:* As programs are not asked to report progress along their specific theories of change, the expert review teams will map reported achievements against the expected sequence of achievements along the elements of the CRP and Flagship theories of change (or alternative program logic models). With that, the review team will be able describe what and where progress has been made toward reaching stated objectives and link these to learnings about the theory to change and influencing factors. The benefit of this approach is that it describes the program progress toward objectives more clearly than counts or lists of deliverables, providing a better understanding of (a)

the plausibility of cause-effect linkages within the program logic and (b) the contribution of the CRP to development outcomes.

- *In-depth analysis of selected outcomes and impacts*. The expert review team will select one or two Outcome and Impact Case Reports (OICRs) for each CRP, in consultation with CRP leadership. The review will analyse the selected OICR(s) in greater depth, looking at the contribution of the CRP's research in successfully addressing a given development objective, mapping the reported outcome or impact within the Theory of Change at the Programme and Flagship level. This work will be done through analysis of documents from the CRP and from next users of the research, such as national government policies, and interviews with key informants (both within the CRP and equally importantly the next users of the research, e.g., external stakeholders in NARS and national policy-makers) who may assist in better understanding the nature and importance of the CRP's contribution, as reported in the OICR. A specific reporting template for the OICRs analysis will be provided to the review team.
- *Contextual analysis*. For many reasons related to context within the program or the context of those who would move the research forward to development and scale up, research for development projects and programs may progress at a different pace. At a minimum, context of a program will be characterized by the age of the program including all earlier phases of similar research, total amount of budget, quality of funding, and the CRP's typology as a Global Integrating Program or Agri-food System Program.
- *Analysis of management and governance*. There are several sections in the Annual Reports in which CRPs report aspects related to learning lessons as the research evolved and challenges that arose and how those were handled. The annual POWB discusses changes, if any, in the theories of change. The review team will supplement these sources with responses from surveys, interviews and focus group discussions. The analysis will triangulate information from these sources to identify how the CRP has managed and governed its research program in the context of the challenges faced over the period of review.

Deliverables and consultation for the CRP Review

The review team is expected to produce the following deliverables:

1. A preliminary findings matrix, for discussion midway through the review process, to check the progress of the review and to provide a basis for early course correction if required. The CAS Secretariat will provide the review team with a template for the preliminary findings matrix.
2. A brief presentation of preliminary findings, for the debrief with the CRP management and the CAS Secretariat for validation, factual corrections, and feedback.
3. A draft report of the CRP review, for review by the CRP management and the CAS Secretariat for final feedback. The CAS Secretariat will provide a template for the draft and final reports.
4. A final report of the CRP review, following the report template with a maximum of 20 pages, a 2-3 page executive summary, and a set of annexes with additional information apart from the main body of the report.
5. A PowerPoint presentation covering the main points of the review, including purpose, methods, findings, conclusions, recommendations and additional notes relevant to the review. The CAS Secretariat will provide a template for this presentation.

Templates for the preliminary findings matrix, draft and final report, and the presentations will be provided to the review team in the first week of the review.

The review team will engage with the CAS Secretariat and the CRP under review at the following key points:

- Initial discussion with the CAS Secretariat to start the review and clarify questions from the review team;
- Briefing at the start of the review between the review team and CRP management, facilitated by the CAS Secretariat;
- Interview with the CRP Leader and a focus group discussion (FGD) with other members of the CRP management during data collection;

- Debrief presentation of the preliminary findings led by the review team, for validation, clarifications and feedback by the CRP management and the CAS Secretariat;
- The draft report will be shared with the CRP Leader and staff for factual correction and final feedback.

Additional discussions between the review team, the CRP management and the CAS Secretariat may be scheduled based as needed during the course of the review.

Schedule of the reviews

The reviews will be conducted in a phased, stepwise manner, so as to enable due support from CAS Secretariat throughout the review process. The first three reviews will take place between April and June 2020. Thereafter, in late June, CAS Secretariat will conduct an 'after-action review' involving the Program Leaders from the first three CRPs reviewed, for fine-tuning of the review process in enhancing learning and minimizing the burden on CRPs. While refinements to the review process may be made, the fundamental review parameters will remain harmonized for all CRP reviews through the year. Substantive changes on questions and sub-questions are not foreseen from the after-action review. The subsequent nine CRP reviews will be conducted in the second half of the year, commencing in August 2020.

The first set of reviews, scheduled for April through June 2020, includes three CRPs - one global integrated program and two agri-food system programs. This initial selection of CRPs for review is based on (a) two Agri-Food Systems and one Global Integrated Program, (b) the length of time since the last independent evaluation conducted for the CRP and (c) CRPs with and without substantial changes in program and/or structure from Phase I to Phase II. CRPs that had requested to be included in the first set of reviews were prioritized, within the above criteria. The working schedule of CRP reviews is attached as Annex 2. For each review, an indicative time frame of deliverables and milestones for the review is provided in Annex 3.

Qualifications for the expert review team

Each review team is anticipated to include (1) a senior subject matter expert with in-depth subject matter expertise related to the CRP under review, and (2) a senior evaluator with experience in agriculture, natural resources management, food systems or nutrition. Of the two team members, one must serve as the team leader, who will bring relevant experience in that evaluation leadership and be the lead author for the report and accountable for the review team performance.

The estimated number of days of effort for each role in the review is provided below:

- Senior Subject Matter Expert: 40 days;
- Senior Evaluator: 30 days;
- Team Leader (additional to one of the above roles): 10 days.

The qualifications for each role are outlined below. **This is a desk-based review and no travel is envisaged.**

Qualifications for the senior subject matter expert include the following:

- Excellent understanding and knowledge of the key issues in agriculture, natural resources management, food systems and/or nutrition, as related to the CRP to be reviewed;
- 15 or more years (preferably, over 20 years) of work experience in the domain(s) related to the CRP to be reviewed;
- Strong knowledge of the main international institutions and mechanisms involved in the areas of research and development that are the focus of the CRP to be reviewed;
- Academic background relevant to the CRP's areas of research;
- Excellent understanding and knowledge of the international debate on the key issues related to the CRP to be reviewed;
- Depth of knowledge of areas of research and development that are the focus of the CRP to be reviewed;
- Knowledge of the CGIAR and/or the CRP to be reviewed.

- Strong English writing and verbal communication skills.

Qualifications for the senior evaluator include the following:

- 8 or more years of experience leading evaluations, preferably including international programs or research on agriculture, natural resources management, food systems and/or nutrition;
- Extensive experience with theory-based evaluations, including analysis of effectiveness in relation to a Theory of Change with potential implicit adaptations;
- Preference for evaluation experience in one or more research areas specific to the CRP;
- Preference for knowledge of the CGIAR and/or the CRP to be reviewed.
- Strong English writing and verbal communication skills.

In addition, the consultant (from one of the above two positions) who will also serve as Team Leader must demonstrate the following:

- Experience leading evaluation of complex programs, preferably in international agricultural research;
- Demonstrated accountability in terms of timeliness and quality of deliverables and responsiveness in communication;
- Academic background or experience in evaluation and/or an area relevant to the CRP's work;
- Strong project management skills;
- Experience working virtually (online) in successfully conducting interviews and facilitating discussions with senior managers, researchers, practitioners and policy-makers;
- Excellent English writing and verbal communication skills.
- Excellent presentation and report writing skills, including for executive and multicultural audiences and remote/virtual presentations.

Alternative team configurations may be considered, and the CAS Secretariat will discuss options presented by the proposed consultants.

Applications are encouraged from teams of two consultants with the qualifications and experience outlined above for the senior subject matter expert and the senior evaluator, clearly indicating which individual is proposed for the team leader role. Individual consultants may also apply for the subject matter expert or evaluator roles, with the intent to be matched with a suitable counterpart from the roster of other applicants.

Application process

Interested teams and individuals should send their CV and a cover letter indicating the role to which s/he is applying and the CRP(s) in which s/he is qualified to serve as a reviewer (see Annex 1 for an overview of the CRPs and their flagships and also the full CRP profiles on the CGIAR website: https://www.cgiar.org/research/research-portfolio/). The CV and cover letter should include information on the applicant's:

- Proposed role (Senior Subject Matter Expert or Senior Evaluator) and intended CRP(s) for the review, with both the role and intended CRP(s) clearly stated in the subject line of the email and the cover letter;
- Demonstrated expertise in the technical research areas relevant to the CRP to be reviewed;
- Experience in evaluation;
- Expected daily fee rate (demonstrable with evidence of rates on previous assignments);
- Location and time zone of her/his work location;
- Email, telephone and Skype contact details of the applicant(s);
- Names and contact information (email, telephone and postal address) for three (3) referees, who will be contacted for short-listed candidates;

- Availability for the CRP review based on the schedule provided in Annex 2, as well as more generally over the period April through December 2020.
- List of publications (including peer reviewed work and past evaluations/reviews authored)

Interested teams or individuals meeting the above criteria should send their application by email to CAS-Evaluation@cgiar.org . Applications are accepted on a rolling basis, and the CAS Secretariat will contact short-listed candidates for follow-up at an early date, for potential scheduling of the relevant CRP review. Regrettably, we are unable to respond to all applicants, but will retain CVs and contact information on file for those who meet the above criteria.

Contract and payment schedule

The CAS Secretariat is hosted at the Alliance of Bioversity International and the International Center for Tropical Agriculture[9], at the offices in Rome, Italy. Consultancy contracts will be issued by the host institute of the CAS Secretariat. The members of the review team are expected to abide by the Conflict of Interest and Safeguarding policies of the CAS Secretariat and its host institutions, and must maintain independence in fact and appearance from the CRP under review throughout the duration of the assignment. Each review team member must sign and return statements indicating their understanding and compliance with the policies of the CAS Secretariat and its host institutions.

Payments under the contract are scheduled as below:

- 25% on signing of the contract;
- 25% after the midterm check-in discussion and delivery of the preliminary findings matrix, subject to satisfactory approval by the CAS Secretariat;
- 50% on delivery of the final review report, subject to satisfactory approval by the CAS Secretariat.

This is a short-term consulting opportunity with the level of effort as indicated for each consultant role. All consultancy fees and conditions will be administered in line with the Alliance for Bioversity International and CIAT's approved policy for consultants.

Contact at the CAS Secretariat for the CRP 2020 Review

The CAS Secretariat has appointed an Evaluation Consultant, Dr. Ravi M. Ram, to manage the CRP review process, along with CAS evaluation staff and a consultant providing senior technical advice, under the overall direction of the CAS Secretariat Director, Allison Grove Smith. Questions regarding this Terms of Reference should be directed to r.ram@cgiar.org .

Who we are

CGIAR is a global research partnership for a food-secure future. CGIAR science is dedicated to reducing poverty, enhancing food and nutrition security, and improving natural resources and ecosystem services. Its research is carried out by 15 CGIAR Centers in close collaboration with hundreds of partners, including national and regional research institutes, civil society organizations, academia, development organizations, and the private sector. These 15 Centers have close to 10,000 staff based in over 50 countries.

Each Center has its own governing instrument, board of trustees, director general, and staff. CGIAR Research Centers are responsible for hands-on research programs and operations.

The CAS Secretariat supports and facilitates the CGIAR's independent advisory services, comprising the Independent Science for Development Council (ISDC), the Standing Panel on Impact Assessment (SPIA) and an independent evaluation workstream.

9 Bioversity International and CIAT are CGIAR Research Centers. For further information consult the websites at https://www.bioversityinternational.org and www.ciat.cgiar.org

In 2020, CGIAR is embarking on an ambitious reform, One CGIAR, to streamline governance and operational structures and processes across CGIAR. More information can be found here[10].

The Alliance of Biodiversity International and CIAT is an equal opportunity employer and strives for diversity

[10] https://www.cgiar.org/how-we-work/strategy/cgiar-system-reference-group/

Annex 2: 2017-2021 CGIAR Research Portfolio

The 2017-2021 CGIAR Research Portfolio consists of two sets of CGIAR Research Programs and a cluster of Research Support Platforms; source: https://cgiar.org/research/research-portfolio/.

Agri-Food Systems Programs

Fish Agri-Food Systems (FISH)

Forests, Trees and Agroforestry (FTA)

Livestock Agri-Food Systems (LIVESTOCK)

CGIAR Research Program on Maize (MAIZE)

Rice (RICE)

Roots, Tubers and Bananas (RTB)

Wheat (WHEAT)

Grain Legumes and Dryland Cereals (GLDC)

Global Integrating Programs

Agriculture for Nutrition and Health (A4NH)

Climate Change, Agriculture and Food Security (CCAFS)

Policies, Institutions, and Markets (PIM), which includes a Gender Collaborative Platform

Water, Land and Ecosystems (WLE)

Research Support Platforms

Genebank (GENEBANK) supports CGIAR genebanks represent the largest and most widely used collections of crop diversity in the world, with 768,576 accessions, including 25,301 *in vitro* accessions and 28,063 *in situ* accessions. Several FPs mention genebanks in their ToC.

Big Data in Agriculture (BIG DATA) aims to mobilize CGIAR data to accelerate research and spur new data-driven innovations, build data collaboration internally and externally, and leverage CGIAR expertise while claiming leadership in digital agriculture. It also supports and promotes Open Data.

Excellence in Breeding (EiB) aims to modernize breeding programs, drawing from the public and private sectors to provide access to cutting-edge tools, services and best practices, application-oriented training and advice to increase the effectiveness and efficiency of breeding. This resulted in some efficiency improvements[11].

[11] Performance Report 2017 Transforming the Global Food System. https://www.cgiar.org/executive-summary/ accessed May 11.

Annex 3: WHEAT Budget and Staff

Data for both budget and staff have been provided by WHEAT.

Annual budget during the review period, in U$S.

Item	W1W2 + Bilateral			W1W2		
	2017	2018	2019[A]	2017	2018	2019 [A]
Management & CapDev	1,575,000	1,361,000	2,260,000	1,575,000	1,361,000	2,260,000
FP1	5,609,000	3,234,000	3,001,000	1,454,000	1,428,000	1,536,000
FP2	10,300,000	7,755,000	7,260,000	2,315,000	2,595,000	2,552,000
FP3	31,078,000	16,880,000	19,715,000	6,662,000	6,913,000	5,311,000
FP4	17,923,000	13,916,000	13,169,000	1,924,000	1,467,000	1,804,000
Partnerships[B]	0	1,840,000	0	0	1,840,000	0
Total	66,485,000	44,986,000	45,405,000	13,930,000	15,604,000	13,463,000

[A] 2019 Includes 130k for ICARDA management which was separated from research funds + CRP management reserve of 534k, both per MC decision.

[B] In 2018, in order to demonstrate the level of investment in external partners, it was decided to separate the partnership grants total, hence this is not shown as part of each FP. In the end of year spending this is reflected in the per FP numbers.

Scientific (SCI), technical (TEC) and administrative (ADM) full-time equivalents in WHEAT flagships for the 2017-2019 period. T:S is the ratio of technical-to-scientific staff and A:P is the ratio of administrative-to-production (scientific plus technical) staff averaged for 2017-2019.

FP	2017			2018			2019			Average 2017-2019			T:S
	SCI	TEC	ADM	SCI	TEC	ADM	SCI	TEC	ADM	SCI	TEC	ADM	
1	12	25	4	12	25	4	12	24	4	12	24	4	2.1
2	13	49	3	13	49	3	15	46	3	14	48	3	3.5
3	40	336	37	37	340	35	37	343	36	38	340	36	8.9
4	19	128	12	17	132	11	18	134	11	18	131	11	7.2
WHEAT	84	538	56	80	545	52	82	547	53	82	543	54	6.6

Annex 4: Documents reviewed

All documents reviewed are cited in the main text and list of References. We used hyperlinks for key documents.

Annex 5: Interview list

Name	Affiliation
Michael Baum	WHEAT CRP
Hans-Joachim Braun	WHEAT CRP
Daniel Calderini	Universidad Austral
Mariano Cossani	South Australian R&D Institute
Mina Devkota	WHEAT CRP
Olaf Erenstein	WHEAT CRP
Tony Fischer	CSIRO
Shaylyn Gaffney	WHEAT CRP
Bruno Gérard	WHEAT CRP
Sarah Hearne	WHEAT CRP
Eric Huttner	ACIAR
Victor Kommerell	WHEAT CRP
Marta Monjardino	CSIRO
Rachid Moussadek	WHEAT CRP
Vinay Nangia	WHEAT CRP
Kevin Pixley	WHEAT CRP
John Porter	University of Copenhagen
Barbara Rischkowsky	WHEAT CRP
Ravi Singh	WHEAT CRP
Gustavo Slafer	Universitat de Lleida
Richard Trethowan	CAIGE, University of Sydney
Jacques Wery	WHEAT CRP

Annex 6: Bibliographic databases

Here we present features and references to the two bibliometric data bases.

Web of Science (WoS) is a publisher-independent global citation database. Its multidisciplinary platform connects regional, specialty, data and patent indexes to the Web of Science Core Collection across disciplines and time from over 1.7 billion cited references from over 159 million records. Web of Science is part of Clarivate AnalyticsTM. Further details: https://clarivate.com/webofsciencegroup/solutions/web-of-science/

The SCImago Journal & Country Rank (SJCR) is a publicly available portal that includes the journals and country scientific indicators developed from the information in the Scopus® database (Elsevier B.V.). Citation data are drawn from over 34 100 titles from more than 5 000 international publishers and country performance metrics from 239 countries. SCImago is a research group from the Consejo Superior de Investigaciones Científicas (CSIC), University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares, dedicated to information analysis, representation and retrieval. Further details: https://www.scimagojr.com/index.php.

Annex 7: Survey on quality of science

Survey for WHEAT Flagship Programs, Review 2017-19

Aim

The aims of this survey are (i) to ensure the review captures the perspectives of FP leadership, and (ii) fill gaps or complement information from other sources. Emphasis is placed on Quality of Science.

Background

The review focus is 2017-2019, although we understand the continuity of research and the lag-time from research to adoption of technologies.

The survey comprises four sections addressing the review questions:

1. To what extent does the CRP benefit from sufficient high-quality **inputs** (with reference to the research environment and project designs)?
2. To what extent do the CRP **management processes** ensure the quality of science, including credibility, legitimacy, relevance to next stage users, and potential effectiveness, of the research and operations?
3. In what ways are the research **outputs**, such as germplasm, knowledge tools and publications, of high quality?
4. To what extent is the CRP positioned to be effective in the **future**, seen from the perspectives of scientists and of the end users of research (such as policymakers, practitioners or market actors)?

If you feel that some questions are better answered verbally or need clarifications, please contact me by WhatsApp (+61 428 100 275) or we can organise Zoom/Skype.

Please email this back to victor.sadras@sa.gov.au by 26 May

Flagship:

1. INPUTS

1.1. Please rate the suitability of resources available to meet outputs and outcomes, in a scale from 1 (unsuitable) to 5 (suitable):

Infrastructure:

Data management:

Scientific staff:

Technical staff:

Level and predictability of funding:

Other (please specify):

1.2. What is the single most limiting resource (briefly explain)?

1.3. Beyond resources, are there other limiting factors (e.g. politics, time frame, forced to study certain items)?

2. MANAGEMENT PROCESS

The review will consider the WHEAT periodic re-assessment of the demand and quality of research; the research work environment as enabling QoR4D; research ethics, transparency and procedures for conflict of interest; and use of learning mechanisms to inform current and future research, for ultimate users of the research.

2.1. How is research planned, documented and monitored?

2.2. How do you manage research ethics, transparency and conflict of interest?

2.3. How do you manage the tension between competition and collaboration?

2.4. How do you manage the tension between formal evaluation and priority-setting procedures, and fostering scientific creativity?

3. OUTPUTS

The review will assess external recognition of WHEAT outputs as high quality; collaboration for innovation with next stage users and/or beneficiaries; value of outputs in developing capacities for researchers, next stage users and partners.

3.1. Who are the main end users of your FP research outputs?

3.2. Please rank the importance of RD&E outputs in your FP from 1, less important to 6, most important:

Knowledge:

Scientific publications:

Extension material (specify formats if necessary):

Technological outputs (varieties, practices; other please specify):

Policy outputs (please specify/provide example):

Capacity building (eg postgrads, farmer training; other please specify):

3.3. What metrics do you use to quantify quality of science, credibility, legitimacy, relevance potential effectiveness; and in a scale from 1 (poor) to 5 (very good) how do you rate your FP for these parameters.

Parameter	Metrics	Score (1-5)
Quality of science		
Credibility		
Legitimacy		
Relevance to next stage users		
Potential effectiveness of outputs		

3.4. Legitimacy

The review team has a gap in information regarding Legitimacy. We would appreciate if you can elaborate and provide material/statistics relevant to Legitimacy for the period 2017-2019. Information of networks of institutions and countries is particularly valuable.

3.5. What are the two more important management practices to achieve targeted volume and quality of outputs?

4. THE FUTURE OF WHEAT AND YOUR FP

The review will assess the readiness for adoption of the program's deliverables at the IDO and sub-IDO levels; and changes in the program's enabling environment, capacities and partnerships that prepare its research outputs for successful use by next users and beneficiaries.

4.1. What is the readiness for adoption of the IDOs and sub-IDOs of your flagship?

4.2. To what extent is your FP positioned to be effective in the future, from the perspectives of both scientists and end users of research?

4.3. What is the programmatic evidence for future effectiveness within the life of the program (through 2021), considering the comparative advantages of the CRP and its Flagship Programs and drawing on the CRP's and its Flagship Programs' progression according to their Theories of Change?

5. OTHER COMMENTS

Annex 8: Survey on effectiveness

Semi-structured Interview
Topic Guide

Location/Organization			
Interviewer(s)		**Time**	
Interviewee(s)		**Role in Organization**	
Date of interview		**Location of interview/Type of interview (phone, Skype, in person, group)**	**SKYPE or ZOOM**

I. **Introduction**

- **(Establish rapport)** My name is Donna Podems. I am working for the CGIAR. Thank you for making the time to meet with me.
- **(State purpose)** I am here today to ask some questions about **X with a focus on the period 2017-2019**. *We will use this information to help understand more about the intervention and its successes and challenges.*
    - **(Consent)** This interview is entirely voluntary and you may choose not to participate. If you agree to participate, you can choose to stop at any time or to skip any questions that you do not want to answer. Your interview responses are confidential, and will only be shared among two team members, for analysis. While we aim to use information and perspectives that you provide, in the report we will not link these to your name. Your name will only be listed as a person that we interviewed, in the report annex. Please feel free to stop this interview at any time.
    - **(Transition)** Do you have any questions for me before we start?
    - **Note oral consent given: YES NO**

I would like to ask 3 specific questions about X.

1. **Describe the key achievements for the 2017-2019 period. (See last page of this document)** The last page (page 3) contains the **X Summary of achievements**. If there are additional, significant achievements not mentioned, **please add them to the summary,** using a hyperlink to provide detailed information, if *possible* (e.g. the article, the conference).
2. **What was achieved that was not planned, if anything? Who valued this?** (If these achievements are already described in the X Summary attached, please highlight them, so donors and others who read the report are aware of unplanned achievements). (Did these all fit within your TOC?)
3. **What did X plan to achieve that did not happen? (**What kept it from happening? What can be done to support the result to happen? Please let me know if there is a document that summarizes what was planned but not achieved, as I can summarize or append that document to the final report.)

I would like to shift your attention to how WHEAT is managed and ask 2 questions.

4. **In the time that is left for WHEAT, what would you change, if anything, about the management (e.g. funding, coordination, communication, reporting, M&E), so that you can better implement your intervention?**
5. **There is a Theory of Change for FP2. How do you use the ToC, if at all?**

I would like to end the interview with two final questions.

6. **We spoke about a few topics. The CGIAR and donors will read this report. Knowing this, what do you think should be emphasized in the report?**

7. **Please list any people whose insights are critical for this review. Please keep in mind the key focus is on did X achieve what it aimed to achieve (accountability).**

INSTRUCTIONS: Please make any necessary corrections or changes. At the bottom, please provide any other significant or important achievements that should be included. Please note that we are working with extremely tight page limitations, thus I need to work with short, concise (and hyper-linked) information.

(SHORT SUMMARY PROVIDED FOR REVIEW PRIOR TO INTERVIEW)

Thank you. Please submit your information to donna@otherwise.co.za at least one hour prior to the agreed interview time

Annex 9: Quality assurance

The review management team provided the review team with access to information and resources in four sections: Bibliometric Data, CRP background documents such as POWB and AR, Dashboard Data and Production (review team).

Here we outline how we defined the suitability and reliability of these resources, with one example in Quality of Science and one in Effectiveness.

Quality of science – ensuring reliable sample for bibliometric analysis

The original bibliography files (publication title, authors, journal, and bibliometric data) were incomplete (2019 was missing) and truncated. Questions were raised about the reliability of these data. As an alternative, the review team sought original data from WHEAT. The reviewers found duplications in the database of WHEAT, which were largely related to multiple records as papers are first published online and then recorded again when journal volume and page number are assigned. After a quality control process, we consolidated a reliable database with 469 publications.

In parallel, CAS assembled a database returning 263 papers for the same period. Jointly with CAS colleagues, the review team investigated the gap and concluded that this was partially the result of CAS using Web of Science as a source. Our analysis shows that Web of Science has an incomplete coverage of WHEAT publications (Fig. 2B in main body of report). Importantly, we observe that Web of Science coverage is not only incomplete, but also biased because it does not include non-indexed journals by definition, and systematically overlooks lower-impact journals. Any analysis based on CAS database would therefore over-estimate the impact of publications by WHEAT.

Effectiveness – the risk of uncritical use of highly-aggregated data

CGIAR dashboard is a powerful tool. However, reviewers have to take the outputs of the dashboard at face value. The outputs are highly aggregated, the reliability of raw data to reach these outputs is unknown, and context for interpretation is not straightforward. Bar-charts reporting milestone achievement are an example of all these issues (Annex 12). These issues were investigated by the review team with input from CAS and WHEAT. The main observation from this inquiry was that CAS uses categories to represent milestone achievements that are not the same categories used by WHEAT. The practical conclusion was that CAS data were unsuitable for the assessment of milestone achievement (see Section 2.2.1).

In summary, the review team was critical of data and analytical process, and paid particular attention to the suitability and reliability of data. Highly-aggregated data were used as a guide, and primary sources were the core material for our conclusions. For example, Annual Reviews pointed out some achievements, which were not taken at face value but traced to their original sources (Section 2.2.1). In most cases, primary sources supported conclusions from highly-aggregated sources. In other cases, close inspection of primary sources did not fully support them, e.g. the conclusion that wheat-rye crosses have delivered one line with high resistance to the Russian wheat aphid had to be placed in the context that experiments were conducted under controlled conditions; this important finding was only possible by critical assessment of original data (Section 2.2.1). Similarly, our quantitative analysis of bibliography was triangulated with the reading of each and all papers in a subsample (Section 2.1.3.2).

Annex 10: Limitations

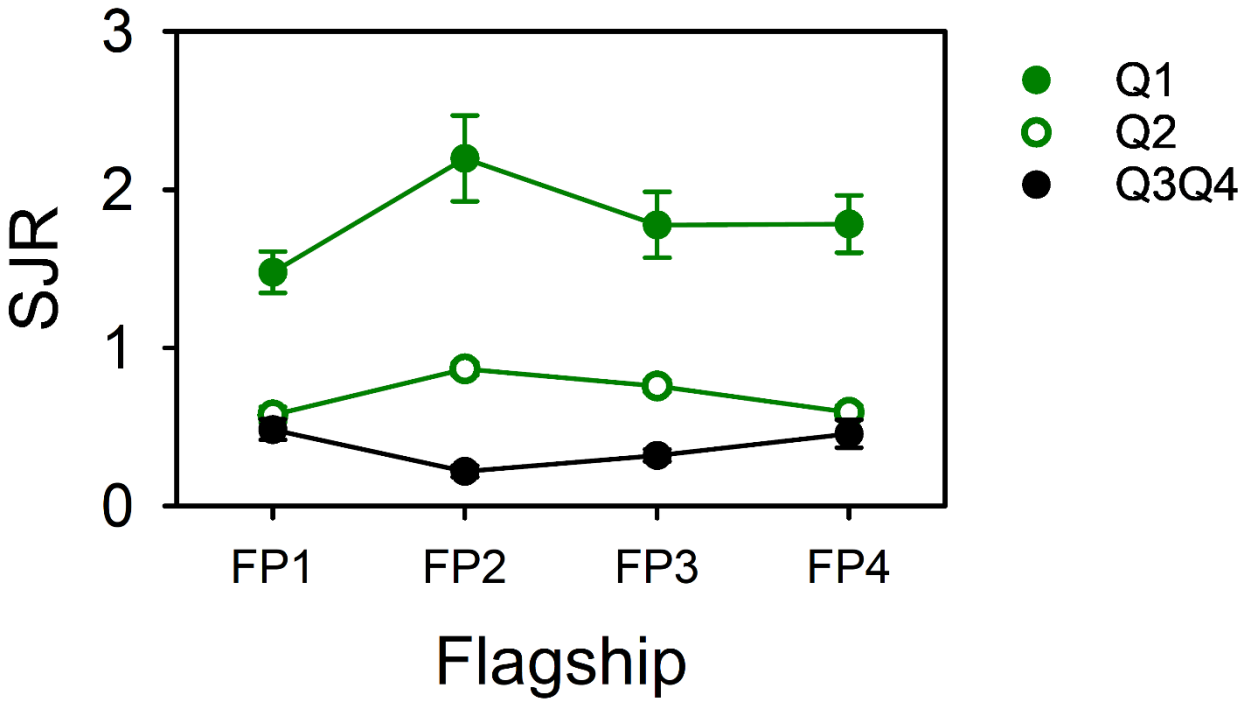
Research limitations and challenges		Mitigation strategy
Lack of an inception report	A planning phase did not exist. Refining evaluation questions, clarifying methods, identifying data sources, and refining a report outline, did not take place.	Refined questions and a refined outline were submitted to the CGIAR though not accepted.
Lack of a valuing framework or agreed process to explicitly value the findings	Research and evaluation are different, and one key difference is that evaluation values something. The evaluation lacked a clear valuing framework or agreed valuing process.	None.
Primary reliance on secondary data.	We are reliant on a considerable amount of secondary quantitative data, of which we have no control over its quality.	We used primary and secondary qualitative data to triangulate all data, where possible.
Contradictions arising from diverse data sources.	Diversity of data sources and types may lead to difficulties in synthesising the findings so that they emerge into a coherent narrative. The research process encountered significant contradictions in the analysis, particularly between the desk-based analysis of secondary data and the primary data collected by the evaluation team.	We used a structured and systematic approach to the analysis, triangulation and synthesis of the data which helped resolve contradictions in the analysis or provided a transparent means of explaining why the contradictions occur.
Unclear evaluation questions/evaluation questions changing mid-process.	Unclear evaluation questions that contained words that needed further defining, which were then revised yet not defined approximately midway through the evaluation process.	Core questions were addressed based on the outline, which provided guidance that implied intended use and information needs.
No contact with primary evaluation users.	Lack of interaction with primary users, despite the CGIAR assigning a UFE approach, limited the evaluators ability to ensure evaluation use. This also limited selection of OICRs as it was not clear what specific outcomes or themes were of most interest to the intended users.	Suggested that evaluation team have an opportunity to engage with the primary users. (Did not take place).
Contradictions between data sources and inconsistent reporting compromising statistical analyses.	For example, the 2018 database for training included "Mexican", "MEXICANA", "Mexico" as entries for nationality; the original bibliographic data set was a single year with truncated data;	None.

	graphs of network analysis used outdated data. The terms of reference of the review feature redundancies and omissions; for example, the purpose does not mention quality of science (Annex 1) and the identified evaluation approaches not entirely appropriate.	

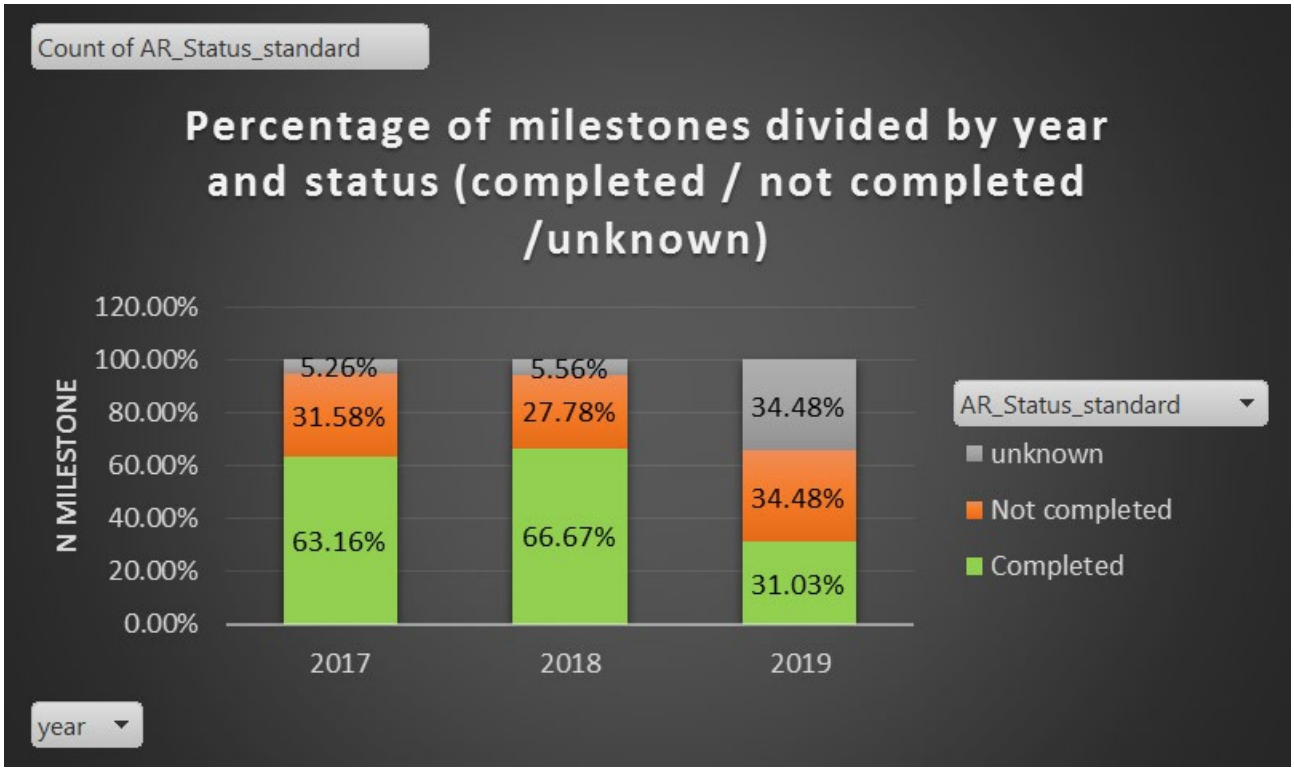
Annex 11: Test of bias in SJR

The impact of a journal and its classification in quartiles depends on discipline as illustrated in Table 1. Hence, we tested if the SJR for a given quartile varied between flagships. Analysis of variance showed SJR varied with quartile as expected ($P<0.0001$) but did not vary with flagship ($P>0.81$) or flagship-by-quartile interaction ($P > 0.82$).

The figure shows SJR average ± s.e., highlighting the similarity in SJR across flagships for a given quartile.



Annex 12: Examples of data in CGIAR database



Annex 13: OICR selection

Selection of two OICR. Although a researcher or evaluator could study an entire population, it is often not feasible, cost-effective, or even necessary to do so. Thus, evaluators select a sample (with each person, place, or thing selected called a case) from the identified population, in order to provide data that answer the evaluation question. An evaluator using sampling needs to provide a transparent strategy that clarifies the criteria used to select the sample and recognizes the limitations. Qualitative sampling is fundamentally different from quantitative sampling. Quantitative researchers and evaluators use probability and nonprobability strategies to select samples to study. A probability sample utilizes some form of random selection. In qualitative inquiry, the approach used by the Evaluation Team, evaluators select specific people, places, or items within the larger population because of the unique insight and rich information they bring to bear on a particular evaluation question. Here the focus is on effectiveness. This section provides the transparent reasoning used to select two OICRs for the evaluation.

What is a case? In research, and evaluation, there are different understandings of what constitutes a case. Based on interviews with WHEAT and the CGIAR, the evaluation team has concluded that an OICR constitutes a case for WHEAT. In qualitative research, a case is often considered as a stand-alone product—it provides a detailed and rich story about a person, organization, or program. Thus, taking a "deep-dive" into a case that should already be a deep-dive, presents a bit of a conundrum.

Boundary setting. The OICRs place a boundary around some phenomenon of interest, and the placement of the boundary is critical. The boundary setting process determines what the case is and therefore the focus of the inquiry. WHEAT has defined that boundary when it constructed the OICR, the unit of our analysis.

Sampling framework. Interview data suggest that the most successful examples in WHEAT are put forth to become an OICR; thus the evaluation team is selecting from a sampling framework that only contains successful cases (which, logically, are cases that demonstrate effective implementation). Therefore, what will be explored "in-depth" for each CRP, and for WHEAT, are two examples of selected from the identified best cases that demonstrate effectiveness, or high-impact cases.

Sampling approach. The sampling frame consists of high-impact cases, and the purpose of selecting two cases is to demonstrate WHEAT's effectiveness in the evaluation report. We chose the utilization-focused sampling approach to guide our selection. The UFE sampling approach selects a set of cases (in this instance, 2) that focus on an issue where sufficient depth and detail will support rigorously identifying key factors that can credibly inform future decision making. In this case, the broad theme is effectiveness.

Sampling criteria. The evaluation team used five sampling criteria: (1) one case on breeding (2) one case on agronomy; (3) access to key informants; (4) cases that have a maturity level 2 or 3; and (5) the likely usefulness to the broader agricultural sector. Given the evaluation's broad scope, other criteria could not be identified. For example, if CGIAR had asked us to identify specific factors that made WHEAT effective, such as partnerships, gender or other cross-cutting issue, strategic approaches, use of W1 or W2 funding, or influencing policy, additional criteria would have formed the selection process. An additional OICR was added (OICR 2764) as it overlapped with the OICR selected, and as such added depth to the narrative.

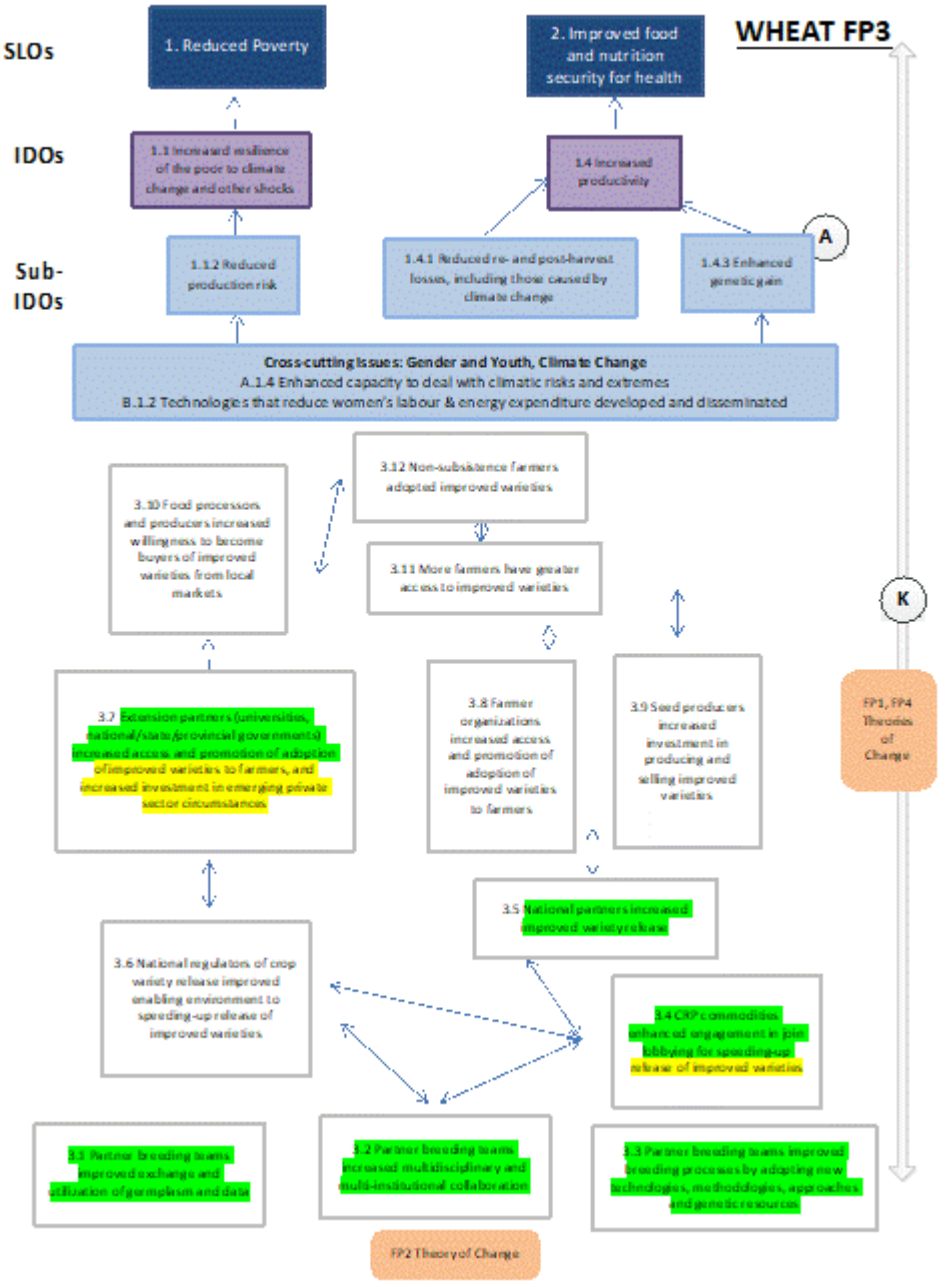
Using the criteria above, the three OICRS selected are:

1. **Wheat Agronomy**. OICR 2524 and 2764. Conservation Agriculture in South Asia (Maturity Level 2).

Wheat Breeding. OICR 3284. CGIAR synthetic wheat breeding strategy successfully transfers valuable diversity from goat grass to modern wheat, providing farmers with climate-resilient, pest and disease-resistant wheat. (Maturity level 3).

Annex 14: Example of ToC

Below is the FP3 ToC that was marked up, and is meant to be reviewed along with, Section 2.2.4. The ToC example is placed in an annex due to formatting challenges.



Annex 15: Conflict of interest statements by the reviewers

1. Main employer and any other organization that provides you with remuneration (which may be named participants in the project/program/proposal you are being asked to review/evaluate)
   Employer: South Australian R&D Institute, Primary Industries and Regions South Australia
2. Are you aware whether a relative, close friend, close colleague or someone with whom you have financial ties is receiving funding from or giving advice to a project/program/proposal you are being asked to review/evaluate?
   No
3. Does any project/program/proposal you are being asked to review/evaluate cite any of your own current research?
   No
4. Does any project/program/proposal you are being asked to review/evaluate name researchers with whom you have active collaborations, recently published joint papers or are in regular email correspondence?
   No
5. Does any project/program/proposal you are being asked to review/evaluate name any of your past PhD students are active participants?
   No

Declaration: I declare that the information provided on this statement is true and complete.

Name: V.O. Sadras

Signed:

Date: 8/4/20

1. Main employer and any other organization that provides you with remuneration (which may be named participants in the project/program/proposal you are being asked to review/evaluate)
   Employer: OtherWISE: Research and Evaluation
2. Are you aware whether a relative, close friend, close colleague or someone with whom you have financial ties is receiving funding from or giving advice to a project/program/proposal you are being asked to review/evaluate?
   No
3. Does any project/program/proposal you are being asked to review/evaluate cite any of your own current research?
   No
4. Does any project/program/proposal you are being asked to review/evaluate name researchers with whom you have active collaborations, recently published joint papers or are in regular email correspondence?
   No
5. Does any project/program/proposal you are being asked to review/evaluate name any of your past PhD students are active participants?
   No

Declaration: I declare that the information provided on this statement is true and complete.

Name: D. Podems

Signed:

Date: 30/4/20

CGIAR Advisory Services (CAS) Secretariat

Via dei Tre Denari, 472/a, Maccarese (Fiumicino), Italy

tel: (39) 06 61181 - email: cas@cgiar.org

https://cas.cgiar.org/