

## ISPC Assessment of the Platform on Big Data revised proposal (2017-2022)

### ISPC PLATFORM RATING<sup>1</sup>: A

#### 1. Summary

- The social value of the data and knowledge products generated by CGIAR is arguably comparable to that of the content of the genebanks. This strongly suggests that CGIAR has dramatically underinvested in the curation and maintenance of data. This is the time to seize the opportunity to capitalize on these assets, since recent developments in linked open data and food systems ontologies are creating new and favorable conditions for achieving the objectives of this platform. The field is changing so fast that the only way to stay on the edge is to be invested and involved in these processes, which largely occur outside CGIAR.
- This is a strong proposal for a platform addressing a crucial and long-standing weakness. The proposed platform is an important means of improving CGIAR system level performance with good potential to generate System-level benefits. The original budget proposal was judged as inadequate to support the desired aims and did not reflect the importance of the topic. As suggested by the ISPC, the revised proposal keeps a base budget and has added an alternative “uplift” scenario with a budget 83% higher than originally proposed, broken down in a modular manner to help prioritization under a limited funding scenario.
- CGIAR and its partners have generated a rich and complex mix of multi-location, multi-disciplinary data and associated information. There is wide recognition that more could, and must, be done to ensure that these data are made accessible for sharing, interrogation, or repurposing – and that this would represent a significant IPG. Furthermore, the nature of CGIAR’s research is so data-driven and data-intensive, that a coherent and strategically positioned coordinating platform on Big Data and ICT is essential and timely.
- In the guidance for pre-proposals for Phase II of the CRPs, it was noted that a number of scientific organizations have already invested in data capabilities and infrastructure. This creates an opportunity for CGIAR to leverage this investment to advance the global public good mission of CGIAR, in coordination and alignment with these international stakeholders. The platform aims to increase the impact of agricultural development by embracing Big Data and ICT approaches to solve research for development problems faster, better and at greater scale. As outlined in the SRF, this will initially be across CGIAR, but is extensible to agriculture at large.
- The Theory of Change of the platform focuses on increasing the capacity of CGIAR and partners to embrace Big Data and ICT approaches. The platform’s strategy focuses on collaboration between CRPs and centers, leveraging external expertise to enable unrestricted discoverability of linked open datasets. Through the “INSPIRE” module, Big Data pilot projects will be launched.
- The proponents of the Big Data platform have responded positively to the ISPC suggestions for strengthening the original proposal, and have satisfactorily addressed all concerns raised.

<sup>1</sup> A+: Outstanding - of the highest quality, at the forefront of research in the field (fully evolved, exceeds expectations; recommended unconditionally).

A: Excellent – high quality research and a strongly compelling proposal that is at an advanced stage of evolution as a CRP, with strong leadership which can be relied on to continue making improvements.

A-: Very good – a sound and compelling proposal displaying high quality research and drawing on established areas of strength, which could benefit from a more forward-looking vision.

B+: Good – a sound research proposal but one which is largely framed by ‘business as usual’ and is deficient in some key aspects of a CRP that can contribute to System-wide SLOs.

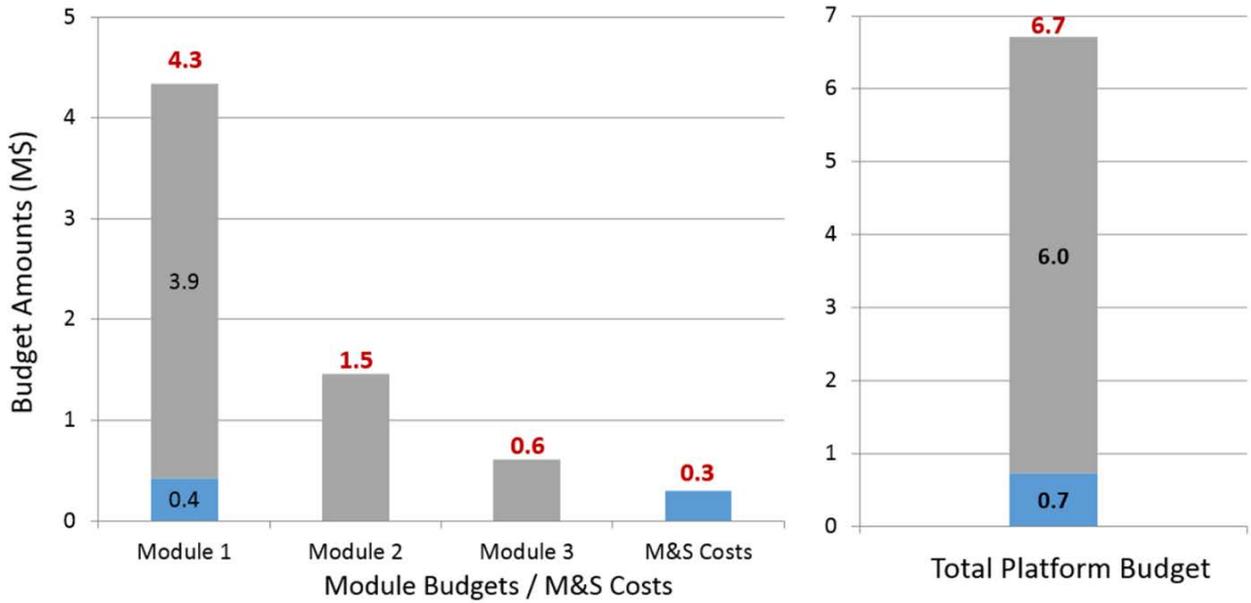
B: Fair – Elements of a sound proposal but has one or more serious flaws rendering it uncompetitive; not recommended without significant change.

C: Unsatisfactory – Does not make an effective case for the significance or quality of the proposed research.

## Big Data 2017 Module and Platform Budgets: Secured W1/W2 Amounts and Shortfalls (US\$M)

- 2017 Budget Amounts (all from W1/W2) not yet Secured
- 2017 W1/W2 Amounts Considered Secured

Figures in red are Total 2017 Budgets Needed



Data Source: CGIAR System Management Office

## 2. Assessment of the Platform response to the ISPC major comments

Initial ISPC comment (16 June 2016)	Platform response/changes proposed (31 July)	ISPC assessment (14 September)
<p>1. The ISPC suggests that the proponents provide a revised proposal that includes development of <b>two budget scenarios</b>: (1) the current baseline budget; and (2) a significantly higher budget, with associated elaboration of activities for each.</p>	<p>Specific adjustments were made to the baseline budget to address ISPC suggestions (adding funding for a core computer scientist, secretariat operations, international board functions, etc).</p> <p>The proposal now includes a significant “uplift scenario” budget for secretariat personnel, supplies, and services; to double the INSPIRE project and impact assessment funding; private sector collaboration on farm management decision-support systems; use of cloud-based infrastructure; data science and curation; ontologies; high-throughput phenotyping analytics; and a drone imagery platform.</p>	<p><b>Satisfactorily addressed.</b></p> <p>The Big Data team deserve praise both for their realistic approach to budgeting as well as for their response to the ISPC suggestion.</p> <p>The additions to the baseline budget required to respond to the specific ISPC suggestions resulted in an increase of 0.77%.</p> <p>The uplift budget (83% above the base budget) is a significant increase in R4D and other platform activities, which is clearly spelled out in the addendum. If even a sub-set of these uplift activities can be funded, this would be worthwhile, and the modular approach to most of the uplift is a practical compromise between the funding in hand and aspirations to achieve full potential of the platform.</p>
<p>2. Fully realizing the opportunities of the Big Data Platform requires <b>additional computer / data scientists</b> as part of the core staff team and hence additional budget for professional staff.</p>	<p>Additional computer scientist included in the base budget as a core member of Module 2 (CONVENE).</p>	<p><b>Satisfactorily addressed.</b></p>

Initial ISPC comment (16 June 2016)	Platform response/changes proposed (31 July)	ISPC assessment (14 September)
<p>3. A section is needed to address <b>ethical considerations</b> in relation to the requirement for IRB approval before data are published.</p>	<p>New annex 8 (one page) and costs for a community of practice added in the uplift budget.</p>	<p><b>Satisfactorily addressed.</b></p> <p>The approach in Annex 8 is a satisfactory response to the ISPC concerns regarding this challenging and urgent issue, but arguably the level of ambition should still be higher. The ISPC agrees that ethical issues are likely to grow, but also feel they must be addressed from the outset in the establishment of the platform.</p> <p>The four bullet points elaborating on topics to be considered by the CoP are:</p> <ul style="list-style-type: none"> <li>• the use of genetic resources and traditional knowledge associated with such</li> <li>• research involving animals</li> <li>• participatory research approaches and the promotion of farmer’s rights</li> <li>• biosafety risk assessments</li> </ul> <p>These are all important, but none of them fully address the challenge outlined here. The more the CGIAR moves towards common research sites with trans-disciplinary work, the more the various sectors of research will need to understand that taking human subjects research seriously will require them to adjust sample sizes, adopt different methodologies for study design, and incorporate their engagement with farmers/herders into approval processes that they then have to stick to. While the CoP may work on these issues, the real question is how CGIAR institutions and CRP leaders will be brought on-board to tackle these critical research process issues.</p>
<p>4. Plans for <b>Big Data analytics</b> need to be mapped out more clearly.</p>	<p>Priorities based on the results of a Montpellier workshop have been added to the proposal for selected topics. There is now a greater emphasis on INSPIRE projects (aligned with CRP priorities) as vehicles for data analytics.</p>	<p><b>Satisfactorily addressed.</b></p> <p>The steps are realistic and appropriate. The ISPC also would like to endorse the proponents’ argument that “it is however important not to overly prescribe these topics, but rather ensure that they are demand driven by the AFS and integrative CRPs, and fully embedded in their workstreams.” It seems likely this articulation with CRPs could be a challenge, but we agree it is indispensable and could make or break the platform.</p>

<b>Initial ISPC comment (16 June 2016)</b>	<b>Platform response/changes proposed (31 July)</b>	<b>ISPC assessment (14 September)</b>
<p>5. An outline of the process of <b>curation, quality review and certification</b> of data should be provided (not more than 1 page) in the annex to the revised proposal.</p>	<p>One-page appendix 9 was added along with budget lines to support the activities described.</p>	<p><b>Satisfactorily addressed.</b></p> <p>As with the treatment of ethics in appendix 8, it is difficult to do justice to this complex (and often frustrating) topic in a single page. However, also like ethics, it is important to have these issues documented in the proposal, together with some concrete and highly appropriate steps intended to address these issues. Though the budget has been augmented in line with these activities, the CGIAR has not had a good record in these areas and it is hoped (finally) that sufficient funding will be committed to make significant progress. While modularity in the uplift budget is sensible for most areas (assessment in point 1 above), this is an area of urgent need in which critical mass likely is needed to achieve necessary results, and so should be given high priority in funding decisions.</p>
<p>6. ISPC recommends setting up an <b>Executive Management team</b>, including the platform coordinator (PI) and leaders (co-PIs) of each module.</p>	<p>Added to the proposal (as a short paragraph on p. 10 in Section 1.0.5).</p>	<p><b>Satisfactorily addressed.</b></p>
<p>7. ISPC recommends creating an <b>International Advisory Board</b> with representatives of relevant initiatives around the world.</p>	<p>Added to the proposal (as a paragraph on p. 11 in Section 1.0.5) and the proposed budget has been adjusted for these costs.</p>	<p><b>Satisfactorily addressed.</b></p> <p>Including the intention of involving representatives of the key international initiatives.</p>