

ISPC Commentary on the Expressions of Interest (EOIs) for the CGIAR Coordinating Platform on Big Data and ICT

Summary

Taking a horizontal view of the 5 EOIs submitted and assessing them based on the objectives and the criteria proposed in the guidance for pre-proposals, indicates that 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 – to also influence the design and implementation of CRP II research programs. The understanding and expertise on the importance and potential of big data across CGIAR consortium centers and CRPs varies, but there are some excellent and internationally recognized teams that could be the driving force for such a Coordinating Platform. The comparative analysis of the 5 EOIs reveals that most of them are very well developed with some excellent ideas; and some are demonstrating senior interest. This is a clear indication of the awareness and maturity within the CGIAR to develop and run such a Coordinating Platform on Big data and ICT. There is complementarity in the proposed activities of the various EOIs, and areas of overlap in the activities that the different proposals bring forward. The ISPC believes that by putting together the best ideas described in these five EOIs, a coherent strategy can be delivered for a Coordinating Platform that may efficiently and effectively address challenges and opportunities across the CGIAR research portfolio.

With this in mind (and considering that only one Coordinating Platform can be selected for funding), the **ISPC considers that the EOIs led by CIAT and by IFPRI (respectively I and IV) stand out, and that CIAT (and its named partners) should co-ordinate the development of a full proposal, drawing on the complementary strengths of the IFPRI-led EoI in particular.** The ISPC also notes its expectation that this platform should have a limited lifespan, aligned with the duration of CRP phase-II. It also considers that the other three EOIs proposed have strong and relevant components that should be considered for inclusion in the development of the full proposal.

NB: the five EoIs submitted are reviewed sequentially (without any particular order) and a general conclusion is included at the end of this document.

I. EOI title: “Leveraging CGIAR Data: Bringing big data to agriculture, and agriculture to big data” (lead partner: CIAT)

Overall assessment

This EOI proposes a platform in the wider sense of an engagement and collaboration medium. It aims to enhance CGIAR’s capacity to deliver big data analytics and ICT-focused solutions through ambitious partnerships with initiatives and organizations outside the CGIAR, promoting CGIAR-wide collaboration, and developing new partnership models with big data players. It particularly aims to organize and promote collaboration internally within the CGIAR; to increase visibility of CGIAR big data analytics and ICT-based initiatives through new channels and collaborative spaces; and to deploy some high profile, integrative, and collaborative big data analytics projects around core themes in the CGIAR CRPs and SRF.

The proposal includes a novel way in which big data may be further communicated to the CGIAR ecosystem: it proposes activities that can raise awareness and showcase the potential of good big data projects (by implementing a number of flagship sub-projects that it calls *Inspire*) but also supporting incubated big data experiments that will be attracted and selected through an open process (by a number of pilots that are called part of the *Venture Capital*). The budgeting of the project is connected to the number of Inspire and Venture Capital sub-projects to be supported every year, which gives flexibility in terms of requested funding. The management structure is light and includes a PI team from CIAT, two Co-PIs from other CGIAR centers, and two from external partners. The proposer team has a long, credible and convincing track record.

Overall, this is an innovative proposal that focuses on activities that are flexible, dynamic and can create significant impact, including comprehensive genomics, field and systems level farming practices. The partnership is very strong but also open to the participation of all CGIAR centers, the EOI seems to offer possibilities for a truly consultative, collaborative, and inclusive approach. The link to the CGIAR OA/OD activities is very important, in order to avoid overlaps of effort. On the other hand, there is a lack of horizontal technology development activity that will put in place some core, backbone infrastructure services – such as a data set registry, a data publication and sharing mechanism, etc. And the proposed EOI does not cover all expected objectives and outcomes, as they are specified in the call for pre-proposals. There is also little mention of ontologies, controlled vocabularies, and other methods for linked open data (perhaps the team took this for granted?). **The ISPC considers this EOI Satisfactory and recommends inviting the partners to lead the development of a full proposal, after revising the implementation plan and taking into account all the key objectives outlined in the guidance call and some of the complementary components proposed in the IFPRI-led platform in particular, and also (where appropriate) from the other EOI Big data and ICT EoIs.** [Score: A]

1. Excellence and quality of the proposed coordination of Lead Center and partners

The leading team from CIAT has an excellent profile and track record. The confirmed Co-PI from IFPRI also has a very good profile. The teams of the two confirmed external partners (IBM Research and the KSS collaborative) are excellent and highly qualified. A list of additional partners (many of whom have confirmed their contributions) is also provided, with an excellent line up of organizations and experts. The majority of them are US-based, but this is a tentative listing that the lead partner aims to further revise at the full proposal stage. The majority of the

partners have a big data analytics and data science background. The open approach in engaging and involving other CGIAR centers as partners is duly noted.

2. Level of ambition described in the collaboration/network and the commitment of the participants/partners

This is a quite ambitious proposal that aims to significantly raise the awareness of big data across the CGIAR, but also the image of the CGIAR to external big data players. All core partners express their strong commitment in the EOI.

3. Strategy for system wide networking

The proposed set of activities, especially in the Convene cluster, is detailed and appropriate to achieve system wide networking and engagement.

4. Quality and efficiency of the implementation including strategy for strengthening expertise across the system

The implementation plan is well designed and quite flexible, although not serving all targeted objectives of the envisaged Coordinating Platform. Technical development is part of the three independent *Inspire* sub-projects. However, there is a need for allocating resources in developing some core backbone infrastructure and services.

5. Potential impact

The proposal clearly describes links to existing CGIAR initiatives and ways in which it will openly invite and engage researchers and data scientists. It also highlights the complementarity of the proposed platform to the CGIAR OA/OD initiative, in order to support and facilitate the publication, linking and discoverability of research data across CGIAR centers and CRPs. The envisaged activities may have a very high impact since they are organized around continuous interactions with the CGIAR stakeholders and external players.

6. Contribution to establishing and strengthening a durable cooperation between the partners that will contribute to the CRPII Portfolio and the SRF

The proposed ‘platform’ aims to adopt the model of CSI, the robust platform for collaboration in geographic information science that was established by CIAT in 1999, and that for the last ten years has been very ably led by IFPRI. Since its inception, the CSI has functioned effectively despite multiple reforms and complex institutional dynamics. The proposed big data platform will follow similar principles that have proven successful and enduring, such as the collegial culture of data and information sharing among CGIAR’s geospatial community as well as its self-governance approach, which is attractive to participants.

Additional specific comments

This proposal is quite differentiated from the others, in the sense that it does not aim for a single big data and analytics ICT platform, but it rather provides space to different implementation sub-projects. These *Inspire* sub-projects are independent from one another, focus on different case studies, and may be implemented by different partners. They can have differences in terms of the data types and formats that they use, as well as the big data technologies and tools that they will engage. This is quite an interesting and innovative approach, considering that many of the big data technology providers that would like to support CGIAR services and applications work on solutions that may not be easily combined (or are not always open source). Giving them the flexibility to work on different technologies and analytic engines is also giving space to the

different CGIAR teams to choose the big data technology provider of their own preference. On the other hand, such an approach somewhat departs from the original concept described in the guidance call. There are parts of a common infrastructure that will need to be put in place – such as the services for data publication, linking, and discovery across the various CGIAR software systems and data repositories. The EOI does not foresee this kind of horizontal data infrastructure development, which the various Inspire sub-projects would also need, if they wish to make their data and services discoverable in commonly defined and interoperable ways.

The proposed mechanism of organizing competitions for small pilots from centers and CRPs that would be financed as small-scale demonstrators and experiments is particularly appealing. This kind of incubation activity can also provide space for low-cost experimentation (and failure) with various big data problems and technologies, creating a pool of ideas that can be then further developed at a more mature level. Again, such kind of small pilots can be easily developed in a disconnected and not-interoperable manner. The existence of an underlying infrastructure will put some structure into the way that they are developed.

The workplan needs revision in order to include horizontal activities that would help develop such a data and semantics backbone: this should include work on better understanding the complexities of achieving interoperability among different data types, models and software systems, in order to allow for each sub-project to register and map the used (or required) data models, to publish them as linked vocabularies or ontologies, to link from one data model to the other, to request transformation of data types from one model to the other, and other similar tasks. Furthermore, this would be the backbone to interconnect and make accessible existing data repositories and systems from CGIAR centers and CRPs, so that the *Inspire* and *Venture Capital* sub-projects can discover and use them. Such a global agro-informatics infrastructure does not need to be a CGIAR-developed or -maintained one: the platform could take advantage of (and further improve and evolve) existing components and tools that are part of a globally shared data infrastructure under the umbrella of CIARD¹ (such as FAO's VEST registry² for cataloguing metadata schemas, extending AgriVIVO/agriProfiles³, etc.). The lead Centers could possibly bring on board partners that have such a view in designing large-scale, distributed and interoperable systems for heterogeneous data linking and aggregation. This would also help the partnership to plan and budget more realistically the work required for such a backbone platform.

The approach in creating links and collaborations with external stakeholders and networks is quite dynamic and crowdsourced, giving the freedom and flexibility to the sub-projects to propose their own partners. This approach seems appropriate but there are some strategic liaisons and partnerships that should be horizontally planned. For instance, work on the Global Agricultural Concept Scheme (GACS⁴), FAO's AIMS⁵, CIARD¹, the Interest Group on Agricultural Data (IGAD⁶) of the Research Data Alliance (RDA⁷), and the Global Open Data in Agriculture and Nutrition (GODAN⁸) initiative are of high importance and relevance.

¹ <http://www.ciard.info>

² <http://aims.fao.org/vest-registry>

³ <http://www.agriprofiles.net>

⁴ <http://aims.fao.org/activity/blog/global-agricultural-concept-scheme-gacs-beta-10-released>

⁵ <http://aims.fao.org/>

⁶ <https://rd-alliance.org/groups/agriculture-data-interest-group-igad.html>

⁷ <https://rd-alliance.org>

⁸ <http://godan.info>

II. EOI title: “Socio-ecological Informatics Platform for Global Agri-food Systems Sustainability” (lead partner: ICARDA)

Overall assessment

The proposed platform aims to start from the assessment of the need for a global agro-informatics platform by focusing on identifying relevant data sets (many of which can be classified as big data ones), used or needed software tools, and appropriate use cases. These are expected to lead to the development of a library of use cases that will be organized and linked with the associated data sets and software tools. And they will be shared and discussed within a CRP-wide community of practice (COP). The EOI has a food land/use system focus and builds on components of the Dryland Systems and other CRPs. The main data types and sources to be examined are those needed as input in order to make interventions that will help address global challenges, and to synthesize outputs from CRPs.

The core partners (ICARDA and CIMMYT, Bioversity and other CGIAR centers) are owners of data sets and catalogues but also stakeholders developing tools on top of such data (such as climate and ecological forecasting, land use visualization tools, crop yield forecasting tools, etc.). An existing COP would be further extended and expanded, mainly involving the integrated systems analysis and modeling of CRP-Dryland Systems, the Geoinformatics Unit at ICARDA, the Agro-geoinformatics group at CIMMYT, and the Development Impact Unit at Bioversity. The governance model is impractical and expects to be open to the engagement of other CRPs. The proposal originally integrates the Dryland Systems, DCLAS, WHEAT and MAIZE CRPs.

Overall, this EOI is well-designed and involving strong teams. The planned partnership is narrow and needs confirmation (e.g. IBM, Google); it could include additional partners that have experience in such platforms and that can also bring in the skillset required for the interoperability and linked data parts. It would have been better to start from a small set of specific use cases, based on the experience of the partners, before going for a bottom-up and open approach. The task of developing the underlying platform seems to be not covered and budgeted satisfactorily in the EOI pre-proposal. **The ISPC considers that the EOI is Unsatisfactory, although it has some relevant components; it does not recommend inviting the proposers to submit a full proposal.** [Score: D]

1. Excellence and quality of the proposed coordination of Lead Center and partners

The proposed teams from ICARDA and CIMMYT have good profiles. Other partners include IIASA, USDA, Utrecht University, NASA, Open University, USGS, JRC, and FAO. Private partners considered include ESRI, IBM and Google. The expected partners still need to be confirmed and would bring on board data sets and forecasting services. There is a lack of expertise around agricultural data infrastructures and agricultural data models, interoperability and linked data. There is also need for a link with the AgMIP⁹ initiative.

⁹ <http://www.agmip.org>

2. Level of ambition described in the collaboration/network and the commitment of the participants/partners

This is a very ambitious project which would require much more than the planned contribution of the core research team members from the three Centers engaged (ICARDA, CIMMYT and Bioversity).

3. Strategy for system wide networking

The decentralized approach to engaging CRPs and giving space for their initiatives is a strong selling point of this EoI. This could be foreseen in the work plan as a dedicated activity for bringing on board CRP use cases that are not represented in the partnership.

4. Quality and efficiency of the implementation including strategy for strengthening expertise across the system

The core work plan activities (in terms of WPs) and the allocated budget are not sufficient to deliver the type of expected and planned outcomes. A complete re-organization would be needed, to also foresee activities related to the development of the platform and the data interoperability/linking work.

5. Potential impact

The proposed integrating platform has strong links to the communities and initiatives working on geo-informatics and climate/ecological modelling. It may be that additional collaborations and partnerships are sought, in order to be able to achieve the desired impact to these audiences and application areas. Global impact on the larger-scale challenges will not be easy to achieve through the currently planned platform.

6. Contribution to establishing and strengthening a durable cooperation between the partners that will contribute to the CRPII Portfolio and the SRF

There are strong connections to a few CRPs that can serve as a solid cooperation basis. A clear mechanism for establishing a durable and sustainable collaboration would be required, such as the mechanism for getting on board use cases from other CRPs. A more focused mechanism like a big data advocacy group that would connect to the CRPs would be more appropriate and efficient, compared to a global consortium scheme that is not achievable with the given resources.

Additional comments

The proposed EOI has a valid argument: existing big-data systems are general services rather than purpose driven that target information on drivers, risks, framing variables, performances of agri-food systems. The main concern is that the partnership currently lacks a more horizontal perspective on the purpose and services of the underlying agro-informatics platform, as an infrastructure that can serve the needs that the proposers rightly bring forward. However the solution may not be “a central informatics system”. To be able to encompass biophysical and socio-economic dimensions with respect to agri-food systems, study their drivers and socio-ecological responses, link data across scales, and efficiently couple data with complementary systems models/tools, an open, modular and interoperable data infrastructure will be needed.

Aiming for a Global Co-learning Consortium for Big Data Agro-informatics (GCCAI) sounds very challenging and ambitious. Proposers would need to narrow down the scope of such an effort into building an advocacy working group for co-learning activities on big data in

agriculture and nutrition. Such a group could also connect to, engage, and bring together existing Communities of Practice within the CRPs - but also other networks.

III. EOI title: “Realizing Digital Agriculture” (lead partner: ICRISAT)

Overall assessment

This EoI focuses on leveraging data in ways that may accelerate the development and delivery of superior crop, livestock and fiber production technology, knowledge and services to farmers in the developing world in real time. It identifies the lack of a shared ‘supercomputing facility’ required for processing and analyzing vast and complex data sets within the CGIAR. It proposes the development of a coordinated platform that will be able to pool such data in a standardized manner and that will try to leverage state-of-the-art analysis protocols and methods from within and outside, to maximize the CGIAR’s collective impact in supporting smallholder farmers. The comparative strength of this EOI is the leadership of Dr. David Bergvinson, ICRISAT’s Director General, and his experience and network from developing the Digital Agriculture Strategy for the Bill & Melinda Gates Foundation.

The main CGIAR partner is ICRISAT, with contributions from several other Centers. Two core partners come from the private sector and have a prestigious profile; i.e., ESRI.com and aWhere.com. Additional secured partners include mostly Indian partners with a good profile (including the Indian Agricultural Statistics Research Institute and Digital Green). An intention to reach out to other governmental and private sector partners is also stated. The proposal invests heavily to the professional network of Dr. Bergvinson and some existing ICRISAT collaborations (such as *One Agriculture One Science*). It aims to take advantage of the GODAN initiative, although there are currently no visible links of the partnership to this network. The main envisaged dissemination channels include MOOCs, involvement in citizen science projects, and an annual ICT and Big Data forum.

This EOI has an excellent lead partner and private sector partners, but with a lack of positioning to the CGIAR internal and external ecosystem. The focus seems to be on hi-tech tools and ICT services, but less on "social software" of how this would span the system. Infrastructure and services plans are not clear, but seem to be more centrally driven than responsive to the system. The planned partnership should have included additional partners that would (a) contribute to the specification and implementation of robust case studies across the domains of coverage and the countries involved; and (b) also bring in the skillset required for catching up with work already done on interoperability and linked data for agricultural research information and outcomes. **The ISPC considers that the EOI is Unsatisfactory, although it has some relevant components; it does not recommend inviting the proposers to submit a full proposal.** [Score: D]

1. Excellence and quality of the proposed coordination of Lead Center and partners

The PI from ICRISAT is its DG with an excellent profile, who would contribute about 20% of his time during the first year and about 10% in the following ones. A full time Coordinator is planned to be hired by ICRISAT. The rest of the CGIAR partners have excellent personnel listed as well, but with a small percentage of their time to be devoted to the project (10%). The private partners (ESRI, aWhere) have excellent data-related profiles and portfolios, with applications to agriculture and previous collaborations with CGIAR centers. The Indian partners from the IT sector in Hyderabad are highly experienced but their contributions to the work are not very clearly described. A listing of additional partners that are going to be invited is provided, but it is not clear which parts of the work they will handle. There is an apparent lack of expertise around agricultural data e-infrastructures and agricultural data models, interoperability and linked data.

2. Level of ambition described in the collaboration/network and the commitment of the participants/partners

This is a quite ambitious EOI proposal which the DG of ICRISAT would like to link to a more general Digital Agriculture agenda. From the perspective of ICRISAT it seems to have a strong commitment at DG-level, but to achieve its objectives and have impact across the CGIAR it will require much more than the planned contribution of the core team members from the CGIAR Consortium Centers involved.

3. Strategy for system wide networking

This is the criterion on which this particular EOI suffers. The current version of the proposal does not include a well-articulated strategy, but rather some stand-alone activities that cannot inform and engage the whole CGIAR system.

4. Quality and efficiency of the implementation including strategy for strengthening expertise across the system

The implementation plan is relatively well designed, with some points that could be further improved. The most important shortcoming is the weak design of activities 6 and 7 that cannot deliver the expected outcomes and therefore strengthen the big data and ICT expertise across the CGIAR system. In general the budget estimation and breakdown is on a cost item level and does not help in making a more specific assessment of the resources allocated to each task – for instance, US\$1M for “Commissioned ICT initiatives” is not a sufficient explanation to help judge the value of this expenditure.

5. Potential impact

The personal involvement of ICRISAT DG in the platform could be a possible driver for impact across the CGIAR Centers, but it needs to be supported by a more strategic plan and a set of operational strategies that can serve this purpose. A potential threat is that this becomes a prestigious and important project by ICRISAT with difficulties in engaging and affecting the CRPs and the rest of the Consortium. Furthermore, a clearer connection to partnership outside the CGIAR is needed for better positioning this integrating platform to other international initiatives and platforms.

6. Contribution to establishing and strengthening a durable cooperation between the partners that will contribute to the CRPII Portfolio and the SRF

It was difficult to see in the current version of the proposed work plan, the mechanisms by which a strong and durable cooperation would be established.

Additional specific comments

The proposed EOI would benefit from a better positioning within the existing research information initiatives within the CGIAR. The proposal is not highlighting how relevant some key existing partnerships of ICRISAT can be to this integrated platform – for instance, in One Agriculture One Science partners like MSUglobal have rich experience on topics related to the delivery of data-powered solutions for research and extension. In addition, the involvement of CGIAR Centers in global research fora and networks should be taken advantage of: an important example is the Interest Group on Agricultural Data (IGAD⁸) of the Research Data Alliance (RDA⁷), where CGIAR centers like CIMMYT and Bioversity lead the work on dedicated data types (such as wheat data interoperability).

The tasks devoted to the study and specification of data models and ontologies, for publishing and linking the variety of research data produced within the CGIAR, are well thought through and planned. On the other hand, the current version of the work plan overlooks (and to some extent plans to overlap) a large amount of work that has already taken place in this area, mostly under the umbrella of the CIARD initiative. It seems that the current partnership would benefit from partners that can enhance its starting point in terms of achieving interoperability among different data types, models and software systems. Achieving this objective will require tasks that can take advantage of the current state-of-the-art in agricultural data semantics, interoperability and linked open data (such as the Global Agricultural Concept Scheme – GACS⁴ by FAO, NAL and CABI). And these tasks need to be appropriately designed, budgeted and supported by experienced partners.

The envisaged architecture of the platform is not very clear, but it seems that it is expected to be quite a centralized system hosted and managed by ICRISAT. Another option would be to look at more distributed architectures and decentralised flows of information, as there are several other research information storage and management systems across the CGIAR. The overall architecture should put in place an open, modular and interoperable data e-infrastructure combining and linking multiple other databases, software systems, and online applications – rather than design everything around a singular database or online application. It may be that some expertise related to large-scale, distributed and interoperable systems for heterogeneous data linking and aggregation would enrich this perspective.

The approach of the EOI refers to multiple (mostly real-time) data sources and streams that will be aggregated and visualized using the powerful ArcGIS engine of ESRI. Experience has shown though that the real complexity of such applications (together with all the practical difficulties in defining or enhancing data models, ontologies, data publishing mechanisms, and online applications) is revealed if a number of clearly specified use cases are selected. Having such a set of specific, well-defined use cases will help partners illustrate the full complexity of exploring which data is of relevance, which software tools may be enhanced, and how these local applications may be connected to the backbone big data platform (for supporting data discovery, linking and processing). The current version of the EOI does not explain which agricultural research partners (from the CGIAR Centers and beyond) will work on developing and testing the envisaged demonstrators.

The EOI needs to explore how it may model the semantics and solve interoperability problems in relation to agricultural modeling – typically looking into ways in which input and output data may be stored, managed and exchanged across application areas and software systems. A major initiative to consider is the AgMIP. The envisaged system-wide learning activities can be significantly improved if they are meant to successfully inform the design and implementation of capacity development interventions across the CGIAR. Delivering a series of MOOCs and linking the integrating platform to citizen science project will probably not work effectively in terms of engaging the CRP teams, other relevant CGIAR centers, and inform and support research and extension staff.

IV. EOI title: “CGIAR Big Data Analytics Platform” (lead partner: IFPRI)

Overall assessment

This EOI is proposing the deployment of a CGIAR Big Data and Analytics platform that will be built around two categories of use cases and that will be deployed through an open-access data and analytics portal to serve both CGIAR and external research stakeholders, to enrich the analytical opportunities for CRP research partners. A very strong partnership is envisaged, requesting a quite impressive amount of resources (US\$19M for six years).

The lead CGIAR partner is IFPRI, proposing a solid management structure that will involve some additional CGIAR partners (namely, Bioversity International, CIAT, CIMMYT, ICARDA and IWMI) but will also fund the representation of all Centers and CRPs. An impressive line-up of partners is listed, with more than 25 planned partners with excellent profiles. The role of each partner is well-thought through and positioned in the overall project. The majority of the partners is US-based, which may create some bias in representing the developments and perspectives of other parts of the world. The proposal has a well described plan for the engagement of the CRPs and the collaboration with other CGIAR-funded initiatives (such as the Open Access – Open Data and the Virtual Information Platform ones). It does not have visible links to other regional or international initiatives, such as the Global Open Data in Agriculture and Nutrition (GODAN) initiative. The main envisaged dissemination channels are through the Data and Analytics Community of Practice and the organization of training workshops. The management structure is quite well described in the EOI, although this may still not be sufficient for such a multi-million dollar initiative with so many contributing partners.

Overall, this EOI is very well-written and with a solid justification of its activities. The proposed EOI has a strong IPG focus, is thoughtful on what "big data" is, and on the implications for existing CGIAR datasets with a comprehensive approach to types of data x disciplines. The EOI is strong on ontologies and linked open data (both conceptually and having the right partners) and data quality. The proposed partnership is quite large but of high quality, although this may still benefit from a couple of targeted additions. The work plan is properly designed, although it may still need to justify the magnitude of requested resources and how it would be allocated across activities and partners. There is some overlap with the CIAT-led EOI, but also clear potential complementarities. **The ISPC considers this EOI Satisfactory but recommends that the development of a full proposal is coordinated by the CIAT-led team, drawing also on key complementarities from this proposal.** [Score: A]

1. Excellence and quality of the proposed coordination of Lead Center and partners

A core expert team with senior researchers from IFPRI and the other CGIAR centers has been assembled, who can contribute about 20% of their time as Co-PIs in the project. A full time Platform Coordinator is also expected to be enrolled, although it was not clear whether it is going to be from listed personnel or a new hire. Due to the size of the project, a strong coordination team will be required, including a Research/Scientific Coordinator, a Technology/Platform

Coordinator, and an Administrative Coordinator – not necessarily all from IFPRI. A very interesting and high quality list of partners is presented in the EOI, although some of the proposed partnership may still be a wish-list. The current synthesis and proposed partnership seem to be lacking a partner (or more) that has experience in designing large scale data e-infrastructures and coordinating their implementation and deployment over distributed data sources and streams.

2. Level of ambition described in the collaboration/network and the commitment of the participants/partners

This is a very ambitious project, especially in terms of requested resources. At this level, one would expect to see the commitment of senior personnel responsible for Data and Knowledge Management in the involved CGIAR Centers – and especially in the lead Center. More contribution from the core team members from the CGIAR partners may be required to successfully implement such a large project.

3. Strategy for system wide networking

For such a large initiative (and the amount of funds requested), there may be a need for mechanisms that can push the big data agenda at a very senior level across all CGIAR centers and CRPs. Recruiting part-time Data and Analytics Representatives and organizing a series of training workshops will not be enough. The CGIAR Consortium may need to be represented in the CGIAR Open Access and Open Data initiative, since it is closely linked to the big data platform. A series of strategic measures and mechanisms for higher-level advocacy should be included in the work plan.

4. Quality and efficiency of the implementation including strategy for strengthening expertise across the system

The implementation plan is well designed, but there is need for a revised and more elaborate version of the strategy for strengthening expertise across the system, using a Data and Analytics Community of Practice but also other mechanisms, such as taking advantage of the Capacity Development Community of Practice, creating an advocacy task force that will investigate big data needs and promote the platform throughout the Centers and the CRPs, etc.

5. Potential impact

There is very strong thinking on how this EOI could be relevant to a comprehensive range of IDOs/sub-IDOs (*Annex 2*), with relevance to CRP Theories of Change and impact pathways. The full description of proposed activities of such a complexity and detail can help achieve this goal. Furthermore, a clearer connection to the outside world will help in better positioning this effort to other initiatives and platforms.

6. Contribution to establishing and strengthening a durable cooperation between the partners that will contribute to the CRPII Portfolio and the SRF

The current version of the proposal describes several synergies with other initiatives, within or across CRPs. The Data and Analytics Community of Practice seems like one of the mechanisms that can further strengthen and sustain the collaboration across the system. But an additional series of events and activities is needed that would engage further the CGIAR community and its partners.

Additional comments

The proposed EOI is well positioned in the relevant CGIAR agenda and other initiatives, providing strong potential links with existing initiatives and a well-designed work plan. The partnership is impressive, although it seems to be mostly involving American partners. This seems like a project designed around existing collaborations and partnerships of IFPRI that have proven their value and can deliver the promised outcomes – although this perspective could be somewhat biased and not reflecting the needs of other Centers.

The activities related to the further development of agricultural ontologies are well designed and planned. The involvement of FAO's AIMS team ensures that the integrating platform will build on work that has already taken place in this area. It may be interesting to make these links and liaisons more evident, especially with IGAD of the Research Data Alliance⁷. A semantic backbone that can be used to publish and interlink the various ontologies is still missing. This is a demanding task and the experience of FAO from AGROVOC and GACS⁴ should be exploited. The planned development of specific tools and services for data publishing and discovery is promising, since a list of specific services (like a CGIAR Survey tool) will be delivered. However the platform needs to take advantage of existing components and tools such as FAO's VEST registry² (AgriVIVO/agriProfiles³, GACS etc.).

The description of the envisaged ICT solution and technical architecture is not clear from the current version of the EOI. There are some diagrams that are provided by some large commercial providers, but they are generic and not linked to the existing systems and data repositories of the CGIAR centers and CRPs– not even the pioneer ones of IFPRI. The ISPC strongly recommends a distributed architecture that can manage the multiple decentralized flows of information across the CGIAR centers, and that can serve as a backend, underlying data infrastructure for the agro-informatics platform. This can be decoupled from the front end Data and Analytics portal. This infrastructure should be open, modular and interoperable in order to be able to combine and link multiple data sources, streams, repositories and databases, but also support and serve any software system and online application that will need a strong analytics engine behind. In the current form of the work plan, the use cases are specific and solid but seem to be driven mainly by IFPRI and its core partners. More active involvement of the rest of the CGIAR partners is required for running local pilots in collaboration with the software vendors that already support their research activities. More specific and elaborated measures are needed to explain how the proposed platform will join forces and align efforts with the relevant open data movement and initiatives, such as GODAN and CIARD.

V.

VI. EOI title: “STAR-BID: Supporting Transformative Agricultural Research through Big Information & Data Networks, Strategies and Analytics” (lead partners: ILRI/ICRAF)

Overall assessment

The proposed platform is bringing together two types of complementary activities: i) a Community of Practice, embedded within and across Centers and CRPs, to support their capacity to collect high quality research data that may inform their respective strategies and Theories of Change - and subsequently provide evidence related to progress towards the IDOs and SLOs of the CGIAR SRF; ii) a Network of Data Systems that is envisioned as a distributed network of data systems across CGIAR Centers and CRPs. This will have two main functions: ensure that the diverse datasets of the CRPs are integrated both internally as well as with external agricultural data networks; link the research and data experts working with each data system to guide the analysis of our research across the range of complex agricultural food systems and cross-cutting issues. The platform co-led by ILRI and ICRAF builds on existing activities of the current CGIAR Consortium Data Management Task Force, and aims to strongly align its efforts with the Open Access and Open Data (OA/OD) initiative.

The proposal EOI has an excellent plan for the engagement of CGIAR data managers, the CRPs and the alignment with the OA/OD initiative. It aims to take advantage of existing groups like the CGIAR Consortium Data Management Task Force and the CGIAR Open Access Working Group. And it will put in place a Community of Practice on Data for Impact, which will serve as the mechanism for further engagement and capacity development on data-related topics. The management structure is adequately described, with a full-time manager jointly shared between ILRI and ICRAF to coordinate the project.

Overall, this is an EOI which is focused around the capabilities of the selected partners. The link to the CGIAR OA/OD activities would be a strong asset. The technical approach of the EOI is based on quite traditional solutions focusing on ILRI and ICRAF expertise (joint IT center); but expertise on big data technologies still seems to be lacking. The rough budget estimation and breakdown does not help in making an assessment of the resources requested at this stage, but in general the resources requested for technical e-infrastructure (data centers etc.) seem quite high. The expected activities are not focused enough on acquiring computation resources (“machinery”) and more explanation is needed to justify the value of this expenditure. **The ISPC considers that the EOI is Unsatisfactory, although it has some relevant components; it does not recommend inviting the proposers to submit a full proposal.** [Score: D]

1. Excellence and quality of the proposed coordination of Lead Center and partners

The leading teams from ILRI and ICRAF (two DDGs as Co-PIs) have very good profiles, as well as cross-disciplinary capabilities and expertise. A full time manager is expected to be assigned to the program. The listed partners have very good profiles and the roles assigned to them match their expertise. The proposed platform will be bringing on board a highly skilled and committed group of experts (and a very large number of data sets and data repositories). The CGIAR

Consortium Office is also represented, in order to ensure the alignment with the OA/OD. The main technical partner is the Information Management and eScience group of Manchester University, an excellent unit working on big data processing and analysis workflows for scientific research. Other partners include CABI (providing the link to GODAN), Reading University (as a data curation and quality expert), as well as UNEP and the Kenya Partnership for Development Data (as data providers). Links to the group at the University of Wisconsin that participates in the Knowledge Systems for Sustainability (KSS) collaborative have also been made. This accounts for a rather small and quite balanced partnership, with an emphasis on Kenya and Eastern Africa. However the current partnership does not include groups with proven expertise on big data technologies and analytics (such as Apache's ones and NoSQL databases).

2. Level of ambition described in the collaboration/network and the commitment of the participants/partners

This project puts extreme importance on the data management needs of the researchers and research-support staff. Such a bottom-up approach may not seem too ambitious but could deliver quite efficiently and effectively. The commitment of partners is well documented and solid.

3. Strategy for system wide networking

The project aims to take advantage of the existing data- and interoperability-related task force and working group within CGIAR, therefore the potential for reaching out to data managers across Centers and CRPs is very high. However the project may need to engage senior representatives from CGIAR centers that are not partners in this project. Strategic measures and mechanisms for higher-level advocacy should be included in the work plan.

4. Quality and efficiency of the implementation including strategy for strengthening expertise across the system

There is no clear technical development activity, which explains why the deployment of a specific analytics platform or online services for data publishing, linking and discovery is missing. The part of the work that has to do with the deployment of a Network of Data Systems should be revised and further detailed, including an overall architecture of the distributed e-infrastructure and the specification and development of the various big data components (for data processing, aggregation, indexing, analysis, visualization etc.).

5. Potential impact

The proposal is well coupled with the CGIAR OA/OD initiative that is also expected to support and facilitate the publication, linking and discoverability of research data across CGIAR Centers and CRPs. On the other hand, the coverage that this partnership is expected to achieve would be more in Africa and less in other parts of the world. This could limit the potential impact of the envisaged platform. If the scope of the big data and ICT platform is wider, then the lead partners will need to expand their activities and (most important) the expected partnership to include more pilot sites. Furthermore, a clearer connection to the outside world would help in better positioning this effort to other initiatives and platforms.

6. Contribution to establishing and strengthening a durable cooperation between the partners that will contribute to the CRPII Portfolio and the SRF

Taking advantage of existing cooperation structures in the context of the OA/OD initiative will help establish a sustainable and durable cooperation with other CGIAR Centers and CRPs. The Community of Practice on Data for Impact could be a mechanism to help further strengthen and sustain the collaboration.

Additional specific comments

The proposed EOI is focusing mainly on issues related to data curation, management and publication. It emphasizes developing and sharing good practices in the ways that CGIAR researchers can organize, share and link their research data. It also aims to take advantage of the existing data repositories in both the lead partners (ILRI/ICRAF) and across the CGIAR Centers. Importantly the consortium is very well focused, having well-justified roles for each partner. The geographical coverage of the envisaged case studies is limited, as it seems to be mostly implemented and piloted in Africa – and in particular Kenya and some additional countries to be engaged through regional networks like RUFORUM. The variety of datasets and repositories that ILRI and ICRAF bring on board can help develop diverse and linked use cases; but they are not representative of the needs of other domains and geographical regions.

The tasks devoted to data interoperability and publication as linked data are well described. It will be useful to refine further mechanisms for representing, publishing and linking across data types and models, using the experience that CABI brings from the Global Agricultural Concept Scheme – GACS. Interestingly the EOI adopts a distributed and networking technical approach that will try to build on existing data-related software systems. It is also interesting that the teams at ILRI bring on board a strong e-infrastructure profile that can help them host the platform at their own data centers and test the execution of data processing and analysis workflows over cloud and grid computing - both in house but also (via KENET) through other public research e-infrastructures. On the other hand, the partnership lacks teams that have direct expertise on big data technologies and analytics, especially with applications to agricultural research.

General conclusions & recommendations

Taking a horizontal view of the submitted EOIs and assessing them using the criteria proposed in the report, indicates that 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 – to also influence the design and implementation of CRP II research programs. The understanding and expertise on the importance and potential of big data across CGIAR Centers and CRPs varies, but there are some excellent and internationally recognized teams that could be the driving force of such a Coordinating Platform.

The comparative analysis also reveals some interesting findings that are not obvious while looking at each individual EOI: i) most of the EOIs are well developed with some excellent ideas; and some are demonstrating impressive senior interest. This is a clear indication of awareness and maturity within the CGIAR in order to develop and run such a Coordinating Platform on Big data and ICT; ii) there is potential complementarity in the proposed activities of the various EOIs. Although there are areas of overlap in the activities that the different proposals bring forward; it seems that the different groups are bringing excellent ideas for the implementation of different components of the Coordinating Platform. The ISPC believes that putting together some of the best ideas described in the five EOIs could enable a solid and coherent strategy for a Coordinating Platform that may efficiently and effectively address the challenges and opportunities of Big data an ICT across the CGIAR research portfolio.

Considering that only one Coordinating Platform can be selected for funding, **the ISPC considers that the EOIs led by CIAT and by IFPRI (respectively I and IV) stand out, and that the development of a full proposal should be coordinated by CIAT, but draw in particular on the complementary strengths of the IFPRI-led EoI.** It also considers that the three other EOI proposed have strong and relevant components which could potentially be included in the future platform. The lead partners of the EOIs could work on a common work plan for the Coordinating Platform, where they can be responsible for the clusters of activities where they are demonstrating excellence in the EOIs. This could be codified in an integrating work program that could have a decentralized structure; based on the strengths of the work packages proposed different clusters of activities could designed (e.g. *Requirements & Case Studies, Data Interoperability & Sharing, Agro-informatics Data Infrastructure & Services; Pilots & Demonstrators; System-wide Learning Activities, etc.*).

As recommended by the ISPC strategic study on *Data, metrics and monitoring in CGIAR*¹⁰, the task of putting together a global agri-informatics platform should not be tackled by the CGIAR on its own, considering the significant investments and open infrastructures that have already been developed. It is impressive that technology giants like IBM, Google and ESRI are committing their state-of-art expertise and resources in supporting big data applications of the CGIAR. This is an opportunity to be treated carefully and in an intelligent way by the

¹⁰ [ISPC \(2014\). Data, metrics and monitoring in the CGIAR – A strategic study.](#)

Coordinating Platform: linking the platform to one technology provider or to some legacy software products could limit its further adoption and expansion. This is why the agri-informatics platform would need to be separated into two layers: the Open Data Infrastructure (where open source contributions can be made by all interested public and private stakeholders) and the Big Data & Analytics Services (where different software vendors and solutions can be engaged to solve data-related problems, in a loosely interconnected manner).

The links to the work taking place as part of the CGIAR OA/OD initiative have to be better defined; and this is probably something that should not be tackled at the level of the Centers and an area where this platform could add value. There are several data-powered projects and platforms planned or under development with funding from this initiative, and it would be important to avoid overlaps of work/funding.

Finally, as mentioned in the specific commentaries on EOIs, the full proposal of the Coordinating Platform should seek and establish collaborations with some very relevant international or regional initiatives that are also working on big data for agriculture and nutrition.