**Case Statement for Rice Data Interoperability (RDI) Working Group**

**RDA Rice Data Interoperability Working Group Proposal**



***Photo: IRRI***

TAB Review comments:

Additional comment during TAB discussion of revision:

To achieve full data interoperability there are 3 levels that need to be addressed: technical, semantic/syntactic, and legal. The Rice Data Interoperability WG proposal explicitly recognizes the technical and semantic interoperability issues, but ignores the legal and policy aspects. this interoperability issue should be expressly noted and referenced in the final proposal text, and also state whether the IG plans to address that topic

**1.- Rational**

Rice is a staple food for some 4 billion people worldwide, and it provides 27% of the calories in low- and middle-income countries. Just to keep up with population growth, an additional 104 million tons of (milled) rice beyond the expected 2015 harvest of 475 million tons are needed by 2040, with little scope for easy expansion of agricultural land or irrigation—except for some areas in Africa and South America. Rice farming is associated with poverty in many areas. About 900 million of the world’s poor depend on rice as producers or consumers and, out of these, some 400 million poor and undernourished people are engaged in growing rice.

In the future, given declining environmental quality worldwide, rice will also have to be produced, processed, and marketed in more sustainable and environment-friendly ways, despite the diminishing availability of resources (land, water, labor, and energy). Climate change is exacerbating the situation through the effects of higher temperatures, more frequent droughts and flooding, as well as sea-level rise, which threatens rice production in mega-deltas. Nevertheless, the necessary increases in rice production to meet future demand have to come mainly from increases in yield per unit of land and water. (Rice Agri- Food Systems, IRRI 2015 [www.grisp.net](http://www.grisp.net))

The research and development efforts generated international public goods as well as locally tailored solutions such as publicly accessible data and information systems, genes and markers, breeding lines, improved varieties, improved crop management and postharvest technologies, policy briefs, and training and dissemination materials, as well as knowledge products and capacity building.

The delivery mechanism for these products and services follows a pipeline approach: upstream research results in discoveries and innovations are translated into concrete products, which are introduced, evaluated, improved, and disseminated to intermediate users, and finally become adopted by end users, who may be millions of beneficiaries.

While modern rice research and research data dates back to 19th century, last five decades have seen the successful development of high-throughput technologies that generated large quantities of data in basic, applied and adaptive research in rice sector. However, using these resources comprehensively, taking advantage of the associated cross-disciplinary research opportunities poses a major challenge to both domain scientists and information technologists. Effective data integration and management allows a broader perspective across many disciplines, than is possible from one or a series of individual studies. In the long run, this allows information to be used for purposes other than those for which they were originally intended, to address questions that were unapproachable at the time the data were collected. To this end, the need for umbrella approaches for providing uniform data is a much discussed topic in recent times.

Today global rice research is actually an ensemble of Consortium Research Projects, and is LESS a network of stand-alone institutes. The boundaries between national, regional and global players are blurring. Research organizations need to develop specific strategies and measures to create repositories that quickly communicate with each other.

This warrants building a common framework/ standard for rice data, information sharing. The Research Data Alliance (RDA) through its interest group namely, the Agriculture Data Interoperability Interest group becomes a space to discuss the need to improve data exchange enabling data integration in this domain. A recent success story is adoption of the Wheat Data Interoperability guidelines coming from this interest group by databases from the French National Research Institute (INRA) and URGI (Unité de Recherche Génomique Info). Keeping in view the complexities of data, information, knowledge continuum of rice sector, it becomes imperative to work towards a common framework for rice research data sharing across the globe. Hence we propose the creation of the Rice Research Data Interoperability (RDI) Working Group.

**Context :**

The RDI Working Group aims to reinforce synergies between rice research & development organizations to support food security, nutritional value and safety while taking into account societal demands for sustainable and resilient agricultural production systems.

· provide a forum to facilitate communication between research groups and organisations worldwide on effective sharing of rice research data

· foster communication between the research community, funders and global policy makers at the international level to meet their research and development goals

· facilitate and ensure the rapid exchange of information and know-how among researchers, and support knowledge transfer to breeders and farmers

At the 2012 G-8 Summit, G-8 leaders committed to the New Alliance for Food Security and Nutrition, the next phase of a shared commitment to achieving global food security. As part of this commitment, they agreed to “share relevant agricultural data available from G-8 countries with African partners and convene an international conference on Open Data for Agriculture, to develop options for the establishment of a global platform to make reliable agricultural and related information available to African farmers, researchers and policymakers, taking into account existing agricultural data systems.”

**2. Charter**

The aim of the Rice Research Data Interoperability WG is to provide a framework based on community-accepted standards, which ensure data analysis and data integration facilities. Such a framework is a great asset for the rice community to provide the analysis functions and other services expected by the researchers. Linking databases, platforms and big data from different stakeholder organizations could be helpful for thousands of rice research organizations across the globe especially in Asia and Africa. The Rice Data Interoperability WG in collaboration with partners will work towards bridging the gaps in free data sharing and interoperability of rice research data.

The proposed common framework will help describing, representing linking and publishing rice data with respect to open standards. Such a framework will promote and sustain rice data sharing, reusability and operability. Like the Wheat WG, Rice WG also will try to address questions: which (minimal) metadata to describe rice data? Which vocabularies/ontologies/formats? Which good practices?

With regards to the legal and policy aspects of the underlying data, the proposal will defer to the policies in place in the respective organizations regarding data access: we recognize that private research institutions and companies keep most of their data internal, hence these are never exposed publicly. Nevertheless, private (& for-profit) institutions are encouraged to adopt the proposed framework within their internal systems, with the RDI WG requesting only acknowledgement of the adoption of the Rice RDI framework (and notification to the RDA-RDI WG by email). Most public (or publicly-funded) agricultural research organizations are mandated to provide open access to research data (eg. CGIAR Open Access and Open Data - <http://www.cgiar.org/resources/open-access/>; open data charter for agriculture - <http://opendatacharter.net/introducing-agricultural-open-data-package-beta-version/>), hence all data accessed using the framework are Open Data and should be treated accordingly.

On the matter of interoperability of rice research data, the RDI WG will use the recommendations and outcomes of the Research Data Alliance – CODATA Working Group on Legal Interoperability of Research Data (RDA-CODATA WG) , adapting them accordingly to suit rice research data.

In terms of functionalities and data types, the working group will identify relevant use cases in order to produce a “cookbook” on how to produce “rice data” that are easily shareable, reusable and interoperable. Implementing the framework will help cultivate a rice research ecosystem with people familiar with interoperability, organisations ready to collaborate, and common tools and services. To do so, the WG will focus on;

1. Sharing heterogeneous research data that could be useful across regional boundaries. The rice data may range from germplasm, pedigree, genetics, genotyping, phenotyping, varietal, technological and rice policy data.

2. Capturing the farmers management techniques (tacit knowledge) from varied agro-climatic conditions. Data on Farm innovations can have impact by providing the local solutions to global problems and vice versa

3. Accessing performances of the released varieties. It is estimated that there are about 40,000 varieties of rice cultivated by farmers. The suitability of few varieties in areas other than where they are released can be worked out if only the data related to their performance is shared across the board.

4. Evaluating methodologies to digitize ink-on-paper data (ie, tables of varietal evaluation results), testing 2-4 methods and making the recommendation on the best method(s). There are millions of legacy data generated through multi-location testing programs of National systems in last five decades. The untapped resources of legacy data will bring data revolution in rice sector, if available to thousands of researchers. For example in India, legacy data of 50 years is made available through 27000 datasets related to multi-location trials with effective tagging based on disciplines, year, season catering to data requirements of rice researchers of the country. Digitalization enables access to these datasets by the database interoperability and exchange guidelines to be developed by the working group. We anticipate there will be challenges in automated generation of the correct meta-data associated with the legacy data; we will report these and they could be the target of another WG.

5. Manage the multilingual status of the data: The biggest challenge in agricultural data lies in the variety of languages in which the data is stored in. Notwithstanding the complexities, an attempt can be made in building and piloting a framework that will eventually become a model for agriculture as a whole

6. Bring the International Rice Informatics Consortium into RDA Rice working group: The International Rice Informatics Consortium (IRIC -<http://iric.irri.org> ) aims to provide access to well-organized information about rice, and to facilitate communication and collaboration for rice community, having germplasm diversity as a focal entry point.

**3. Value proposition**

**Individuals, communities, and initiatives that will benefit from the Rice Data Interoperability Guidelines**

The RDI will provide with a linked data framework based on community-accepted standards, which ensure data analysis and data integration facilities. Such a framework is a great asset for the Rice Information System to provide the analysis functions and other services expected by the researchers. Implementing a common framework (however small the scale may be) will help cultivate a rice research ecosystem with people familiar with interoperability, organisations ready to collaborate, and common tools and services.

* The Rice data managers and data scientists will have a common and global framework to describe, document, and structure their rice research data.
* Researchers, growers, breeders, and other data users will have seamless access, use, and reuse to a wide range of Rice data. Data linking will also ease emergence of new data analyses and knowledge discovery methodologies
* Other plants data managers and scientists – will have the benefit of a reusable data framework. Researchers working on other plants will be able to more easily access, reuse and link up rice data with their own data.
* Development professionals and policy makers for taking informed decisions with comparative advantages across the countries
* In terms of scope, more than 132 countries can get direct benefit from the free sharing of rice research data

**Key impacts of the RDA Rice Data Interoperability Guidelines**

* Promote adoption of common standards, vocabularies and best practices for Rice data management. A general awareness among the rice research organizations about openness of data and interoperability standards
* Facilitate access, discovery and reuse of Rice data thereby creating an evidence based impacts for RDA/IGAD/RDI framework
* Facilitate Rice data integration and measuring the impacts of free sharing of rice data
* Creating new opportunities for ontology based knowledge management in rice sector

**4. Engagement with existing work across the globe**

The Rice data interoperability WG is a working group of the RDA IGAD. The working group will take advantage of other RDA’s working group’s outputs. In particular, the working group will be watchful of working groups concerned with metadata, data harmonization, data publishing, as well as the legal and policy aspects of data interoperability.

The working group will also interact with the experts and other projects from national and international organizations and their initiatives which are built on standard technologies for data exchange and representation.

The Rice data interoperability group will exploit existing collaboration mechanisms to get as much as possible stakeholder involvement in the work. The working group will also interact with the Wheat Data Interoperability WG experts and other plant projects such as TransPLANT (http://urgi.versailles.inra.fr/Projects/TransPLANT), agINFRA (http://www.aginfra.eu), GOBII (http://cbsugobii05.tc.cornell.edu/wordpress/) and more generic project such as Elixir Excelerate (https://www.elixir-europe.org/excelerate), International Rice Information System (of IRRI) Integrated Breeding Platform (https://www.integratedbreeding.net) which are built on standard technologies for data exchange and representation, DivSeek and Global Rice Science Partnership (GRISP).

The work will directly align with the ongoing initiatives of hundreds of rice research organizations (including International Rice Research Institute and Africa Rice).

* *Understanding ongoing initiatives - existing work in Rice Data - through a community survey - what difference can WG outputs can make among these initiatives.* The survey will try to understand the systems (data content, ontologies/controlled vocabularies, software technologies including APIs), farmers management techniques, breeding workflow, rice high throughput genotyping/genomics, avenues for harmonizing semantics for phenotyping and agronomy data, ontology based production management and various Interoperability issues.
* *Create a prototype data registry for test in line with IRRI's ongoing work - this helps providing guidelines for creating data registry for rice research organizations.* IRRI and AfricaRice are the lead International Rice Research Institutions that have a reach to every NARS partner. A Data Registry created can immediately be taken to the national partners. RDI WG can leverage the strength of these two organizations.
* *Collect semantics and initiate a framework for a Rice ontology that aligns existing rice ontologies , thesauri, controlled vocabulary and prospect the multi-lingual conversion of ontologies.* Many countries such as India and Thailand build their rice programs on semantic portals using standard rice ontology. RDI WG will work towards aligning all these ontologies to a common framework and design ways to using ontologies for collective intelligence and production and pest management.
* *Best practices for digitization of rice legacy data based on Indian and Thai experiences.* Most valuable and reusable data still lies in paper based documents. While generating the awareness among the policy makers in rice growing countries about the need for digitizing the legacy data, RDI WG will also work towards developing/ documenting best practices for effectively digitizing rice legacy data.

**5. Work plan**

**Form and description of final deliverables**

1. A report on the survey of existing standards among rice research and development organizations. Focus on data availability, accessibility and applicability, formats, ontology, standards and metadata used. A complete analysis of interoperability (or otherwise of) among rice databases and repositories.
2. A set of recommendations on good practices, ontologies, tools and examples to create, manage and share data related to Rice. This work will be based on the existing Wheat Data WG Guidelines[[1]](#footnote-1), The WG will Identify and adopt those relevant to rice data, and will customize accordingly. New types of data might be added according to the results of point 1. The expected output is a Rice Data Framework specification (cookbook)
3. Evaluation of a prototype on Rice specific data registry, Recommendations on how to develop this type of tools would be prepared and disseminated as good practices.
4. Recommendations for a Rice ontology which should align existing rice ontologies, thesauri, controlled vocabularies. This should be the basis for a prospect on multi-lingual conversion of ontologies (TH KU/JP NARO/IRRI/ IIRR / Bioversity) which will not be covered by this WG as a deliverable.
5. Good practices/method(s) for digitization of rice legacy data in line with India's data repository that can serve as a model for getting Thailand national legacy data available and identify best practices (India - IIRR, TH Rice Dept/Ministry of Agric)
6. An adoption phase for deliverables 2 to 5 is foreseen in the workplan, in two different forms:1.) disseminating and creating awareness on the results within the Rice Research Community; 2.) preparing use cases in national and international organizations.

**6. Milestones**

**Month 1 to 6:** Survey to identify the existing standards and recommendations (including vocabularies, ontologies and data formats), end-user categories, and relevant platforms and tools. Target audience are researchers and data managers.

**Month 1 to 12:** First version of the Rice Data Framework specification online (cookbook).

**Month 6 to 18:**  First draft of good practices/methods for digitalization of rice legacy data ready.

**Month 6 to 18:** Evaluation of prototype data registry involving few partners , coordinated by IRRI

**Month 12 to 18 onwards:** Creating general awareness among rice research organizations on common standards, interoperability issues, openness, data standards. This would be followed by an adoption phase targeting one national rice and one international repositories.(organizations to be identified).

**Month 18 onwards:** Creating and measuring the impacts (evidences) of rice data framework and standards of RDI/IGAD/RDA

**7. Adoption plan**

The working group can rely on its initial members to promote a large adoption of the data framework. But the overall aim of the working group is to design a common framework as well as to create the awareness and hence encourage the participation of large number of rice research and development organizations.

The initial work will be undertaken by existing members led by one/ two organizations;

1. Community surveys on ontologies, data needs and interoperability - **Co-lead: IIRR/IRD**
2. Adopt Wheat Data Interoperability Guidelines relevant to Rice Data, and publication of the Rice Data Framework specification (cookbook) –**lead: IRD**
3. Prototype data registry for test and guidelines for creating data registry – **lead: IRRI**
4. Collect semantics and initiate recommendation on a Rice ontology that aligns existing rice ontologies, thesauri, controlled vocabulary and prospect the multi-lingual conversion of ontologies – **co lead: NARO/IRRI/IIRR/Bioversity**
5. Good practices/methods for digitization of rice legacy data - **Co-Lead: IIRR, KU - Thailand**

**8 Initial membership**

**Initial member institutions**

* **IRRI**
* **IRD**
* **Bioversity**
* **NARO**
* **IIRR**
* **CIRAD**
* **CIAT**
* **FAO of the UN**
* **INRA**
* **PhilRice**
* **AfricaRice**

**Initial members**

* Alexandre Guitton
* Devika Madalli
* Imma Subirats Coll
* N Meera Shaik
* Pierre Larmande
* Ramil Mauleon
* Sridhar Gutam
* Manuel Ruiz
* Laurel Cooper
* Elizabeth Arnaud
* Vessela Ensberg
* Giovanna Zappa
* Jeffrey Detras
* Muhammad Naveed Tahir
* Terry Lee
* Vassilis Protonotarios
* Xavier Greg I. Caguiat
* AfricaRice – Ibnou Dieng
1. http://ist.blogs.inra.fr/wdi/ [↑](#footnote-ref-1)