Open Access and Open Data

RICE's information products are international public goods. RICE is committed to their widespread diffusion and use to ultimately achieve a positive, equitable, and lasting impact on the livelihoods of the world's poor rice farmers and consumers. RICE acknowledges the value of its information products, including research data, and considers that widespread sharing of these products will produce scientific, economic, and social benefits. Hence, in line with the CGIAR Open Access and Data Management Policy (OADM Policy) and its Implementation Guidelines, RICE strives to make final versions of its information products openly accessible for use and re-use by others. Information products encompass:

1. Research data: experimental data—on-station and on-farm experiments that generate numerical data plot by plot. These data will be collected either once a year in the wet season or twice a year in wet and dry seasons, respectively; on-farm trail data—participatory trials in breeding that generate numerical data recorded on farmers' fields in addition to information on farmers' varietal choices and farmers' socioeconomic characteristics collected once a year in the wet season; survey data—diagnostic, baseline, and impact assessment studies that generate data on the results obtained from surveys on rice production systems and socioeconomic characteristics of farmers; genomic data—laboratory experiments that generate genotypic data from simple sequence repeats (SSR) or single nucleotide polymorphism (SNP) markers. These molecular data are collected over the year from mapping populations, accessions, and breeding lines; analytical data—data generated by the various laboratories such as those on grain quality; and spatial data—obtained from GIS and remote-sensing analysis and modeling on rice production systems such as maps of rice-growing areas, occurrence of biotic and abiotic stresses, rice suitability, and water resources availability.

2. Publications: peer-reviewed journal articles, reports and other papers, books and book chapters

3. Technologies and tools: video, audio, and images; computer software; web services; novel germplasm products (varieties, pre-breeding lines, discovered genes, QTLs, markers, etc.); novel crop, soil, water, and pest and disease management technologies; technology fact-sheets; guidelines, manuals, and protocols for adoption; varietal passports; training materials; and metadata associated with the information products above.

Planning for and implementing open access and open data: critical issues and anticipated challenges

Most critical issues and challenges relate to ownership, openness and transparency, research ethics, and the FAIR principles: information products should be Findable, Accessible, Interoperable and Re-usable (see section below on Operations). The establishment of ownership is critical to any form of implementation of OADM policies, and is described in Annex 10 on the intellectual asset management strategy. Regarding openness, employees, visiting scientists, consultants, students of the RICE centers are expected to expeditiously

write up and publish scientific data/information products (whether through journal publication, accessible databases, or other means) consistent with the CGIAR OADM policy. RICE encourages publication in open-access journals. Individuals or teams generating data have the first right to publication unless they specifically waive this right. Best efforts shall be used to make all information products open access, subject to the legal right and legitimate interest of stakeholders and third parties, including intellectual property rights, confidentiality, sensitivity, and farmers' rights and privacy, in particular with respect to personally identifiable information. The decision to provide open access will be based on a judgment of the value and relevance of the information product to a wider public and considering the quality and general characteristics of the product and level of processing in the case of data. Data users will be informed that they are not allowed to claim ownership or intellectual property rights over data provided by the RICE centers. Specific conditions are set on the sharing and use of (raw) prepublication scientific data (PSD), defined as all research data, databases, data analyses, data interpretations, draft presentations, reports, manuscripts, intellectual property (whether in preparation or filed, but not published), or other documentation of research results or outputs that are confidential to, or not (yet) disclosed by, the RICE centers and/or their partners. Unless subject to the terms of contractual obligations, all PSD generated by staff of the RICE centers is the property of the RICE centers and subject to the IP policies of those RICE centers.

Activities of the RICE centers are conducted in accordance with the highest ethical standards. The RICE centers are committed to protecting the rights, dignity, health, safety, and privacy of research subjects, and the integrity of the environment when collecting data. Informed consent from study participants will be obtained at the outset of any survey or interview. Personally identifiable information collected with respect to farmers or other stakeholders will be processed fairly and lawfully and, in particular, shall not be made public.

Project planning and implementation

At project planning, agreements will be made among project partners on the sharing of PSD, the anticipated resulting information products, publication strategies, and the storage and sharing media to be used. Where PSD sharing is required to meet the objectives of multiorganization projects or programs, the project team should define principles and procedures for data sharing at the initiation of the project or at an appropriate time thereafter. Such principles and procedures for data sharing may be embodied as clauses in a funding or collaboration agreement and will be assessed by the RICE centers prior to contract execution. Where sensitive data (because of IP, contractual obligation, publication, or other reason) is to be shared between organizations, a confidentiality and nondisclosure agreement will be entered into, which defines the purpose of the data transfer, confidentiality arrangements, and the ways in which the data may be used. Donors or R&D collaborators may request data sharing or confidentiality policies or mechanisms, but such requests must be consistent with the pertinent policies of the RICE centers and in line with the CGIAR Principles on the Management of Intellectual Assets. Within 12 months after completed data curation and quality control, or within 6 months from publication, information products will be made available through dedicated and pertinent open-access media (see next section).

Operations (e.g., technical infrastructure and interoperability considerations, data quality assurance, training activities)

To ensure sustainability beyond the lifespan of CRPs, the RICE centers assume responsibility for maintaining open-access databases and information products, with AfricaRice focusing on the African continent, CIAT on Latin America and the Caribbean, and IRRI on Asia and the world in general. In doing so, these centers will adhere to their OADM policies developed in line with the CGIAR OADM policies and guidelines. The technical OADM infrastructure will follow the FAIR principles described above. Data will be findable and accessible through websites of RICE and the RICE centers and their partners. File formats include jpg, jpeg, xls, csv, doc, avi, mkw, xml, pdf, ascii, and others—preferably open formats that facilitate interoperability. The resources websites of AfricaRice (sections 'publications' and 'rice guide'), CIAT, and IRRI provide access to publications (journal articles, books and book chapters, reports, serials, manual, working papers, research notes, policy briefs, brochures, posters, videos, audio podcasts, images, infographics, and other web tools), data sets (agronomic and socioeconomic, surveys, experimental, statistical, crop, variety, genetic, etc.), and software and tools (e.g., simulation and statistical models, biometric tools, advisory systems, and crop management tools). RICE information products will be made freely available through these and other websites such as dedicated project websites.

Within their pertinent research domains, data and databases are as much as possible interoperable: syntactic interoperability is facilitated through the use of standards such as XML and SQL, while semantic operability is facilitated through protocols such as JSON. Databases can be queried using standard protocols (such as SQL) or through simple web-based and user-friendly interfaces, and are downloadable, accessible, and re-usable through well-described ontologies and explanatory annotations. The publication materials are either fully downloadable, or linked to third-party websites in case copyrights apply (e.g., certain journals that are not open access). For publications that are not copy-righted by third parties (such as certain publishing companies), a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported or a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license is used. Data are only made available that have passed data curation and data quality-control standards (see below under 'workflows and procedures'). Table A9.1 lists open-access databases and information products used to make RICE information available.

The RICE centers will conduct regular training for their staff and partners' staff on data management. IRRI's Risk Management and Quality Assurance-Research Data Management team conducts quarterly training courses that cover various areas of data management, including research data planning; data collection, authentication, and analysis; data storage, backup, and security; and data archival and sharing. The training team also discusses and demonstrates file management tools and software including Beyond Compare, Cobian Backup, WinDirStat, Mendeley, Labfolder, Google Drive, and Dataverse, an online repository for data archiving and sharing also used for RICE data.

Coordination and decision making (e.g., workflows/procedures, governance)

Institutional OADM policies of the RICE centers AfricaRice, CIAT, and IRRI are approved by their center boards and are in line with those of the CGIAR. Responsibility for execution rests with the centers' management teams. The RICE OADM strategy places the centers' policies in a coherent

framework which is overseen by its program management team (PPMT). The RICE centers employ dedicated data management specialists or focal points, who participate in CGIAR's open-access/data management community of practice and who contribute to RICE. For individual projects, project leaders share responsibility in adherence to data management policies by project staff. More complex projects in terms of data type and volume, and type and number of partners, have dedicates data management teams, supported by data management specialists.

The RICE lead center maintains a Risk Management and Quality Assurance (RMQA) committee, which also addresses the management of research data encompassing all aspects from data collection, to curation, quality control, storage, and access. The senior RMQA manager reports directly to the IRRI DG, is overseen by an oversight committee including deputy director generals and directors as members, and is supported by a team of around 35 RMQA officers. The RICE centers have procedures and systems in place for collection, curation, quality control, and safe storage of and access to data. Diligence in adhering to the IA policies by staff of the RICE centers is assessed as part of annual performance appraisals.

Staff of the RICE centers archive their PSD, published, metadata, and other information products on a regular basis in institutional repositories managed by data management specialists. Aggregation of data into databases or other data repositories should occur through processes that clarify publication intent and authorship expectations. The RICE centers use stable, permanent, open-access repositories that enable users and other sites and search engines to access or locate information products. Information products are maintained in central repositories on servers that are protected from environmental hazards such as heat, dust, electrical surges, magnetism, and electrostatic discharges. The RICE centers are committed to making all possible arrangements to ensure the protection of information products against intentional or unintentional loss or destruction. The centers' heads of information, communication and technology play a central role in providing safe data storage and back-up services.

Budget and brief narrative for required resources (e.g., human and financial)

Total budget estimate is around $11 million/year; details are provided in Table A9.2.

Table A9.1. Identification of repository or platform containing information products from CRP projects and/or CoAs and/or FPs for indicative datatypes

Indicative data type	Name	URL	Description
Socioeconomic (FPs 1,2)	World Rice Statistics	http://ricestat.irri.org:8080/wrs2/entrypoint.htm	Yearly national and subnational data on rice statistics (socioeconomic)
Socioeconomic (FPs 1,2)	Farm Household Survey data	http://ricestat.irri.org/fhsd/index.php	Collection of datasets on rice productivity, fertilizer and pesticide use, labor inputs, prices, income, demographics, farm characteristics, and other related data on rice production in farmers' fields. It is a rich collection of actual farm and household data collected through personal farmer interviews, farm record keeping, and periodic monitoring of farm activities from various sites in rice-growing countries of Asia.
Agronomic and socioeconomic (FP 3)	AfricaRice Research data	http://data.africarice.org/	Online repository for archiving and sharing research data; agronomic and socioeconomic data, experimental data, and surveys on rice, in Africa
Agronomic (FP 3)	Dataverse, IRRI	http://irri.org/tools-and-databases/irri-dataverse	Online repository for archiving and sharing research data; agronomic and socioeconomic data, experimental data, and surveys on rice, in Asia
Agronomic and socioeconomic (FP 1,3)	Dataverse, CIAT	https://dataverse.harvard.edu/dataverse.xhtml?alias=CIAT&q=rice	Online repository for archiving and sharing research data; agronomic and socioeconomic data, experimental data, and surveys on rice, in Latin America and the Caribbean
Agronomic (FP 3)	Rice field guides	http://www.africarice.org/warda/guide-field.asp	Collection of field guides for rice crop management in Africa
	WeedSmart		A decision-support tool primarily designed to help farmers manage their fields by providing information that will guide them to efficiently manage and control weeds
Agronomic (FP 3)	WeRise	http://werise.irri.org/	Decision-support system for rainfed rice production

5

Category	Name	Description	URL
Agronomic (FP 3)	Cropmanager	A decision-making tool accessible through the web browser on computers and smart phones, that provides small-scale rice, wheat, and maize farmers with crop and nutrient management advice customized to farming conditions and needs	http://cropmanager.irri.org/
Genetic/genomic Plant breeding (FPs 4,5)	International Rice Information System	IRIS is the rice implementation of the International Crop Information System (ICIS), which is a database system that provides integrated management of global information on genetic resources and crop cultivars. This includes germplasm pedigrees, field evaluations, structural and functional genomic data (including links to external plant databases), and environmental (GIS) data	http://irri.org/tools-and-databases/international-rice-information-system
Genetic/genomic Plant breeding (FPs 4,5) Rice SNP-Seed Database		This site provides genotype, phenotype, and variety information for rice (Oryza sativa L.). The source code is available for partners (IRIC members willing to contribute to IRIC portal development) at BitBucket. Documentation is available at oryzasnp.org/snpseek-javadoc. APIs are available at oryzasnp.org/iric-portal/swagger-ui/index.html. SNP genotyping data (called against Nipponbare reference Os-Nipponbare-Reference-IRGSP-1.0) are from the 3,000 Rice Genomes Project . Phenotype and passport data for the 3,000 rice varieties came from the International Rice Information System (IRIS). The results of the 3,000 genome project were made available at the Amazon Public Storage https://aws.amazon.com/public-datasets/3000-rice-genome/	
Genetic/genomic Plant breeding	Genesys	Global portal for information on plant genetic resources for food and agriculture. Gateway through which germplasm accessions from genebanks around the world can be easily found and ordered. Rice genebank information is included.	https://www.genesys-pgr.org/welcome

Category	Name	URL	Description
Publications	Rice books, IRRI	http://irri.org/resources/publications/books	Complete collection of digitized books published by IRRI. Recent IRRI scientific books on this site follow either a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported or a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license (i.e., Annual Reports). Books that were published before 2007 predate IRRI's creative commons policy. Their "Copyright International Rice Research Institute" statement can be considered to have a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported license.
Publications	Rice publications, IRRI	http://scientific-output.irri.org/	(Peer-reviewed) Journal publications, IRRI
Publications	Rice publications, AfricaRice	http://www.africarice.org/warda/journal-articles.asp	(Peer-reviewed) Journal publications, AfricaRice
Publications	Publications, CIAT	https://ciat.cgiar.org/data-information-knowledge/ciat-research-online	Comprehensive collection of publications, CIAT
Rice management information	Rice Knowledge Bank	http://rkb.irri.org/	Rice production technologies and best rice farming practices, including training materials
Rice reference source	Ricepedia	http://ricepedia.org/	Reference source for rice information
Rice hub knowledge sharing	RiceHub	http://www.ricehub.org/	Virtual platform for Africa's rice hubs, regions where research products and services and local innovations are integrated across the rice value chain to achieve development outcomes and impact
Videos	Rice videos Africa	http://www.africarice.org/warda/guide-video.asp	Rice management instruction videos produced in close collaboration with researchers, field workers, rice farmers, and rice processors. They use simple language and clear visuals, and incorporate lessons from participatory learning and action research. They are ideally suited to build human and institutional capacities in the rice sector in Africa. The videos are available in English, French, and in more than 30 African languages
Videos	Rice videos Africa	https://www.youtube.com/user/irrivideo	Various videos on rice management, the rice sector, and others, Asia

Table A9.2. Budget for open-access and data management in RICE

Human, technical, and other resources	Annual amount (US$)	Explanatory notes
Technology		
Data repository	40,000	AfricaRice: Data Repository; repository upgrading, maintenance, and compliance with the CGIAR core metadata
Publications repository	60,000	AfricaRice: Publications repository; repository implementation, configuration, adaptation, and integration
Hardware/storage (cloud, etc.)	15,000 450,000	AfricaRice: Hardware; computers and office material for OD and OA help-desks
	1,000 workstations at 1,500/year, used 20% for data processing, storage = 300,000	IRRI long-term cloud storage, including archives: $450,000/year (average over 6 years); 100% of which for RICE
		IRRI 1000 workstations used 20% for data processing
Bandwidth	75% of annual bandwidth cost of 600,000 = 450,000	IRRI: 75% use of bandwidth for data transfer
Programming/development/ Website development related to repositories	25 NRS at assistant scientist level = 25 * 18,000 = 450,000	See Table A9.1 for databases and websites programmed/maintained
	10,000	RICE: 25 national staff 100% AfricaRice: Website development related to repositories; develop an entry point to facilitate access, use, and promotion of information products in the data and publications/technologies repositories
Annual maintenance fees	18,000	AfricaRice: Annual maintenance fees: data, publications/technologies repositories hosting and maintenance service pack for 12 month
Other:	50,000	AfricaRice: OA fees for articles; for publications resulting from projects that did not include OA fees in their budget
	10,000	AfricaRice: Marketing and promotion materials

Staffing		
Staff salaries – open data and data management	500 NRS at research level: 0.2 *500 * 13,000 = 1,300,000	RICE: 500 national staff involved 20% in collecting and processing raw data
	125 NRS at assistant scientist level = 100 * 18,000 = 2,250,000	RICE: 125 national staff FTE across RICE involved 100% in curation, storage, archiving, data and managing databases
Data quality/curation	35 NRS at associate scientist level = 0.35 * 35 * 25,000 = 300,000	IRRI: 35 senior national staff involved 35% in overall quality control
	1 IRS scientist level II = 150,000	IRRI: 1 senior international staff involved 35% in overall quality control
<Other>		
Staff salaries – open-access publications and information management	200 IRS level senior scientist I = 0.15 * 200 * 170,000 = 5,100,000	RICE: 200 international staff (scientists) involved 15% in open-access publications and information management
Grouped staff at AfricaRice	20 NRS at assistant scientist level = 0.3 * 20 * 18,000 = 108,000	AfricaRice: 30% time of about 20 staff: Senior Research Data Management Specialist, Senior Technologies and tools Specialist, Senior Publications Specialist, 4 GSS data management staff, 4 GSS technologies and tools staff, 4 publications staff, 5 IT support
Staff salaries – IP/Legal (in support of OA-OD)		These FTEs are capitalized under Intellectual Asset management (annex 10)
Staff salaries – IT (in support of OA-OD), Library	1 IRS scientist level II = 0.3 * 150,000 = 50,000	IRRI: 30% time of CIO
	10 NRS at assistant scientist level = 0.3 * 10 * 18,000 = 54,000	IRRI: 30% time of 10 national staff