View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ICRISAT Open Access Repository

IMPROVING RESEARCH DATA MANAGEMENT AND SHARING: EXPERIENCES FROM ICRISAT

Dr.Guntuku Dileepkumar

Global Leader, Knowledge Sharing and Innovation
Director, Center of Excellence in ICT Innovations for Agriculture
Coordinator, ICRISAT South-South Initiative

Introduction

Since 1971, the CGIAR and its partner organizations have been conducting research programs to contribute to the global efforts on eradication of hunger and poverty through institutional research activities and also through the CGIAR Research Programs. Research data of the activities are very valuable, can be seen as long-term assets of the institutions and can be treated as a major International Public Good (IPG). As these projects are supported with public funds it is essential to extract maximum public benefit from the research data. For the same, several donor agencies are now insisting on data sharing requirements in their grant contracts and for the 'open data' concept to be applied to the research data of the work they fund.

The Department for International Development (DFID) and the European Commission are among those who are adopting this policy. As one of the first custodians of development-related data, the World Bank launched an open data initiative in 2012. The CGIAR followed this example and in March 2012 approved a policy enabling CGIAR research results, including data, be openly available and accessible by default. Any deviation from this policy has to be justified. ICRISAT has initiated its own open access policy in 2009 and was one of the first CGIAR centers with an open access mandate in place. In 2012, ICRISAT developed a data repository for data aggregation, analysis and sharing capabilities in support of the open data movement. The institute approved its data management policy and implementation guidelines in March 2014 and this is in alignment with the CGIAR principles on management of Intellectual Assets, and the CGIAR Open Access Policy.

Open data is not a completely new concept for ICRISAT; in the past it was limited only to specific projects or core programs that have been designed to collect and share data widely.

The role of the scientists especially of those working in CGIAR Research Program mode is changing. Earlier, grant awards and performance evaluations focused on publications in high-impact journals. Long-term data storage and sharing rarely receive allocations in the budgets of research projects. While projects adequately fund data collection and analysis, costs of metadata preparations, the conversion of data into standard formats, curation and data storage and preservation beyond the lifetime of a project are mostly not covered. Data generators should be recognized for publishing high value datasets.

Types of research data

Scientific staff of ICRIAT is collecting, analyzing and synthesizing data on agricultural systems in Asia and sub-Saharan Africa locations, where ICRISAT's research emphasis results in a great variety of research data. This section highlights main data types that ICRISAT research programs produce.

Long term trials

Research programs focusing on breeding of crops need multi-location trials over several years before a new genotype can be fully evaluated. Scientific projects working more generally on combined systems research, natural resource management and climate change also have research spanning over decades in order to characterize and understand changes in these systems. These kinds of data are often collected in various consecutive projects and technical and scientific staff that collect and analyze the data could change.

Baseline data

Baseline data, either household or biophysical surveys, establish the initial condition of a system and can be used, or for an impact assessment of any intervention that results from a given study. Essentially, baseline raw datasets need very strict verification and quality control along with metadata as these are base for future comparison. Usually these data are analyzed and results are presented in donor reports. Baseline data sets combined from multiple projects potentially form high information assets for institutes and the public.

Genomic data

Studies especially on genome sequencing projects, specifically generating the NGS (next generation sequencing) data for gene-phenotype association can quickly generate terabases of sequencing data. One of the key challenges is to devise scalable and robust data management and data sharing solutions. High-performance computing and storage are required to efficiently process the data generated by NGS. Bioinformatics support is integral to address data management systems dealing with efficient storage, retrieval, data mining, data analysis and making data available to the public at the appropriate time.

Data collected as part of a research thesis

Postgraduate students collect a substantial part of the data as part of their research work from the main project. Data can be publically released only after the students have published their papers or thesis.

Value- added secondary datasets

Data collected by another organizations or source or data collected from government publications are considered secondary data; examples are meteorological datasets, remotely sensed data often in the form of satellite images or aerial photographs or panel data sets for rural households. These secondary datasets are shared with scientists from partners under specific user agreements

or licenses and cannot be publically shared by the scientists. Products derived from these data can be shared however, with a reference to the source.

Spatial data

Remote sensing and geographic information system (GIS) data are very important for analysis, interpretation and visualization. These spatial datasets are comparatively big in file size but add value to the research in spatial domain. Spatial datasets are used for analysis, and mapping of large area with less manual intervention. Often these original data are curated and improved and they represent a new international public good that can be described as value-added secondary datasets. Geo referenced datasets can be integrated with socio-economic surveys, soil samples, germplasm collections, and climate data for further statistical analysis.

Data collected in a private public partnership project

While working with private/commercial companies, it is essential to have confidentiality agreements that state explicitly what the data can be used for and which data products can be made publically available.

Partner data

Most of CGIAR Research Programs are undertaken together with partners that sometimes contribute their own non-CGIAR funded data. Here, it is important that the partner decides what data should be used and who should have access to it. Often the partner has agreed to joint copyright of data products, but not the actual datasets.

Barriers to mainstreaming data sharing

Data confidentiality

It would be unethical to make available to the public datasets with personal information. Identifying such data is essential while storage and sharing with the public. Several countries have some form of data protection laws to regulate the processing of information relating to individuals including their traditional knowledge, including the obtaining, holding, use or disclosure of such information. There are a number of standard procedures to anonymize data sets, such as removing names, addresses, and contact information or encryption and hiding of such information. So, it is wise to keep separate datasets for the sensitive and confidential data for internal research use, and for public use.

Data standards and relationship between interdisciplinary data

The multidisciplinary nature of the research conducted within ICRISAT results in a variety of data formats and types. Automated workflows for data verification, cleaning and aggregation need to be customized for each project, resulting in high demand of research support staff time. Interdisciplinary datasets require relational databases with standardized data format that have identified key indicators. To achieve this, metadata formats and survey modules

should be well defined before the data collection that reduce the curation efforts of the scientists. Special care has to be taken during initial data collection and curation which require additional budget and time from the scientists. In addition, ICRISAT has a legacy data of high-value that has not been fully curated and are very important for the adoption or impact studies. To find sufficient budget to update these datasets and to make them available within and outside the center is critical.

Recognition and data ownership

Data cannot be easily protected from copying, or being reproduced without authorization or attribution. Copyright applies not to the facts or the information itself, but to the particular way the facts or information is presented in the dataset or database. As such a database can be protected by copyright, but only in terms of the database model and the data entry and output not the actual numbers or names in the database. The ownership of the data and the right to reproduce the work usually belongs to the institute, and research projects and scientists are given authorship rights. Program directors and project scientists need to ensure that both technical as well as scientific staff are given the deserved credit for their work, thus, all people that have substantially contributed to the creation of the datasets should be data authors.

For projects that consist of multiple datasets produced by different teams of scientists, decisions need to be made on assigning the authorship. Assigning the same authorship (the same persistent identifier) to all project-related data ensures coherence of the datasets, but it does not allow the means to differentiate between different contributions of scientists, which is at times troublesome with respect to accountability. The same issues arise when it comes to dynamic datasets such as R&D databases and trial. Assigning authorship for value-added secondary dataset is more complicated. Based on the contribution definition, data authorship may be provided to these secondary datasets.

Data preservation beyond project life cycle

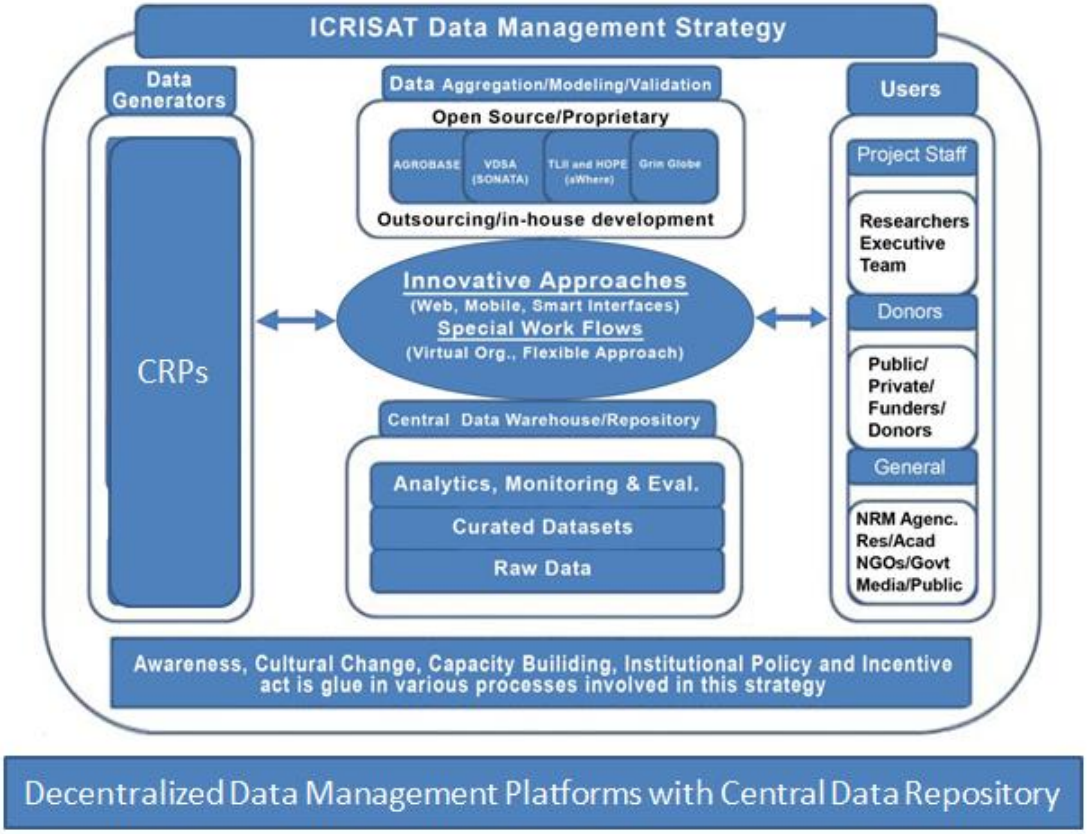
Research projects are usually considered closed when all project requirements have been met, leaving data storage and preservation beyond the immediate lifetime of a project. Metadata collection and proper documentation, especially of multi-country data sets is difficult to outsource and overstretched scientists are often unable to allocate the necessary time. A clear recognition of the value of data in developmental work and a clear mandate to make the data available even after the project closure will ensure that projects will use their scientific expertise to create high-value data sets.

ICRISAT data management strategy

To handle these challenges effectively at ICRISAT, the DMU is working with a strategy to manage the research data that is being produced across the institute, covering all the locations and thematic/program areas; and organize it efficiently in a 'central data repository' by coordinating with research programs

and the Biometrics unit. This will allow the global scientific community "access to the data to get successful results" for addressing pressing issues.

The aim of this strategy is to guide in creation of an environment where institute scientific community is able to produce and share high quality data results, at the same time provide a variety of data management procedures and proven practices at project and program level. This is achieved through decentralized data management platforms specially designed for different data types and program area with central data repository where scientists can use it for data publish. The development and implementation of a data management strategy differs based on the requirements and organizational needs and it is composed of comprehensive metadata for interoperability and controlled vocabulary and ontologies to enable semantic discovery. DMU is periodically in consultation with scientists across the organization, collecting the Program needs and Project requirements.



Goal and objectives

Goal is to developing methods, workflows and protocols to aggregate, store and share the data produced by various research programs.

DMU is working with objectives to make the quality data available to the public based on the institute policies and encourage the staff to use standard methods in curation, metadata conversion and adoption of international standards.

Essential elements of strategy implementation

DMU is working on establishing a clear process for data sharing and management in line with institute data management policy and implementation guidelines. In this process all projects which are generating potential datasets will be identified and a data information specialist to liaise with researcher and data management team to ensure that data is generated, curated and made available through appropriate system. All data that produced in institute should be made available in respective data repositories/platforms and data sharing will be provided based on extent, scope and relevance.

System availability

Establish appropriate data management system for each data type or combination of data type and theme (for instance , aWhere platform for baseline and trial data visualization, VDSA warehouse for survey data). Proper support and clear guidelines have been established and shared to the research staff for better utilization of the system that supports the workflows and protocols.

Cultural change

For a successful implementation of data management policies and principles requires significant cultural shift which is not an easy task. However this change in the attitude of the staff towards better data management requires more efforts and time. Efforts have also been made to bring about a cultural change at the scientists level by continuous interactions and by educating them on the advantages of the data sharing. This is a win-win situation for the institutes as well as the scientific community. Appropriate incentives should be established and metrics on data sharing should be used for performance measurement.

Supporting mechanism

The DMU team has also provided support to staff of various research programs in requirement gathering, and also helped them acquire appropriate proprietary/open source software in support of data management and data analysis.

The DMU does not in any sense replace the stewardship of scientists in matters of data gathering and analysis and the use of such data in publications or training. DMU is a facilitation unit for data preservation and publication rather than for analysis. Under the guidance of the ICRISAT management and in consultation and coordination with the Research Program leadership, DMU is working towards effective ICRISAT research data management with innovative approaches. It has designed and developed several need-based tools for data

visualization and sharing using open source technology and by conducting capacity building workshops to promote open data promotional activities. ICRISAT data management policy is guiding the institute's scientific community on this sensitive and confidential research data and IP, ownership related advises. From technology end, DMU has two powerful servers to host any of the data management applications.

ICRISAT now has many data repositories for different purposes meeting the data management requirement of specific projects, programs and research group. ICRISAT is one of the centers that have all necessary capacities and technologies. DMU is closely working with Consortium data management task force and contributing to the CGIAR.

Mandate on data as research outputs

There should be a clear mandate to publish the research data as research outputs need to be explicitly mentioned in ICRISAT research program framework, which sets common goals, strategic objectives and results to be jointly achieved by the CGIAR Research Program, institute's research programs and their partners. Institute and Research programs have to provide incentives for the scientists to allocate time for data quality and data format at all stages of the research cycle and allocate separate budget for data quality. Assigning a focal point for the project data management activity is also vital.

Working with the consortium office

Informal discussions between data managers led to the realization that many centers face common problems in effective management. After several reviews it was found that there is much common ground between centers in perception of the gaps and shortcomings, as well as a variety of activities at different centers from which others could learn. Recognizing this fact, a CGIAR Consortium Data Management Task Force (DMTF) is being established to expand on institutional and individual member requirements for standardized data management. The task force will be mobilized to undertake duties in defining the priority set of standard vocabularies necessary for describing research products' data and metadata elements in the domains in which CGIAR works. ICRISAT is playing an active role at the consortium level by sharing its experiences and contributing on data repositories and platforms, also bringing the consortium experiences to institute on need based.

Research data infrastructure across the CGIAR centers

In one of the surveys conducted by the Consortium Office, ICRISAT has been identified as one of the centers with proper data management policies and supporting units and repositories. The table below provides a brief summary of the current research data infrastructure across centers. Details for each center are given below. At the time of publication of the report, ICRISAT was working on the research data management policy. Now policy and implementation guidelines have been approved.

(http://library.cgiar.org/bitstream/handle/10947/2884/White_Paper_Shifting_the_goalposts_-_from_high_impact_journals_to_high_impact_data.pdf?sequence=1)

Centre	Research data Management Policy	Data Management Unit	Geo informatics Unit	Biometrics Unit	Centralized Data Archiving & sharing
Africa Rice	YES	YES	YES	YES	Since July 2012
Bioversity	In process	YES			Since Sept. 2013
CIAT	YES	YES	YES		In process
CIFOR	YES				In process
CIMMYT	YES	Recruiting	YES	YES	In process
CIP	YES	YES	YES	YES	YES
ICARDA	In process	YES	YES	YES	In process
ICRAF	YES	YES	YES	YES	Since 2011
ICRISAT	**YES**	**YES**	**YES**	**YES**	**YES**
IFPRI	YES	Recruiting	No		Since 2005
IITA	In process	In process	YES	YES	In process (Partial shared)
ILRI	YES	YES	YES	YES	servers, data partial in development
IRRI	(Currently being updated)	YES	YES	YES	In process
IWMI	YES	YES	YES		YES
World Fish	YES	YES	YES		YES

Additional details can be found at Inventory of CGIAR Consortium Members' Open Access and Data Management Policy

Open data promotional activities at ICRISAT

Open access week

To encourage and educate academic and research communities on the potential of Open Access to Data – ICRISAT organized an Open Access Week observance on 24 October 2013 attended by more than 50 participants ranging from scientists and managers to officers, scholars, interns and partners. "Open Access Booth" was inaugurated on this occasion to allow all scientists and research scholars to deposit a pre/post-print publications and datasets to promote open access movement and this booth was open for two days. This initiative led to a series of requests for uploading almost 100 documents to the open access repositories, which include research articles, publications, presentations, and policy briefings.

Open access and data management policy at ICRISTAT

In 2009, ICRISAT adopted its institutional open access policy http://oar.icrisat.org/mandate.html and setup an open access repository defined

at institute level. In March 2014, ICRISAT approved its data management policy and implementation guidelines to mainstream better research data management, for quality data sharing and keeping it open for public use. This policy is to promote data management practices across the institute and help bring in the required cultural shift in sharing research data and keeping it open access for future use. The purpose of this policy is to promote research data management, curation, preservation, and data sharing across the institute as well as with the global community, while maintaining the stewardship of scientists for data generation, analysis and reporting.

The policy lays out the basic principles of research data management at ICRISAT level, addresses what data can be made available to the public when, how and by who. Roles and responsibilities of individual scientists, projects and research support units have been clearly mentioned and it provided clear guidance on legal clarity on licenses, data ownership and authorships. The policy is accompanied with flexible implementation guidelines to ensure that data is used by the institute and its partners in the most efficient way while seeing to it that scientists are given sufficient time to produce the scientific publications they set out to do.

Capacity building activities

ICRISAT believes in capacity building and its impacts. Hence, DMU at Knowledge Sharing and Innovation (KSI) in coordination with the Biometrics unit organized capacity building programs on "Orientation to innovative agriculture data management platforms" and "Appropriate Technologies and Innovative Approaches for Agriculture Knowledge Sharing" in December 2013. Training on several open access platforms and tools for data management and sharing was provided and scientists and research staff were sensitized on open data sharing and its importance along with ICRISAT Data Management Policy and Implementation Guidelines. On various occasions' brain storming sessions were conducted on data curation, open data, metadata and open data tools.

Technology infrastructure establishment

Within a short period of time, DMU has established two powerful servers with virtualization. These servers are being used for training programs, in-house application development, testing and hosting. By virtualization of the servers, many applications can be hosted with different Operating systems. Maintenance is easy and requires less resource support.

Data loss prevention initiative at individual researcher level

Scientific raw data is very valuable and most vulnerable to permanent data loss primarily due to the sudden crash of the hard drives/ memory devices. The ICRISAT- Risk Management Team has identified the risk of research data loss at institute level in research projects and stressed on the preventive mechanism. Based on the recommendations from the research management committee (RMC) steps have already been taken by the KSI team to protect institute-level data. In

cases of individual-level data loss scientists have approached to KSI requesting help to recover the data from their collapsed hard drives. Considering this, now the ICRISAT-RMT has recommended developing a mechanism to protect the loss of data at an individual scientist level. Based on the needs and recommendations ICRISAT KSI conducted a study on how to prevent valuable research data loss with affordable secured solution and identified Google Drive – a cloud based data storage platform with reasonably good storage space for free (15 GB) that supports wide range of file formats with more user friendly and widely accepted by many people. The team has come up with guidelines to protect the individual level data. Scientific staff are encouraged to follow the guidelines and protect their important research data from hardware failures.

Monitoring & evaluation platform for CRPs

The design and development of an M&E system, based on that operational in the CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), for the ICRISAT led CGIAR Research Programs has been initiated to enable effective monitoring of the activities and outcomes. Two separate databases have been identified for efficient administration and monitoring of R4D: (1) a program management database that maintains project information and budget allocation and utilization of the CGIAR Research Programs that ICRISAT is engaged with, at different levels (Product Line, Cluster of activities, activities) (2) a program information database/website that captures and maintains the identified datasets.

Analysis and visualization tool for AgMIP climate and crop impacts

The DMU team has designed and developed a web platform and desktop application for data aggregation, visualization for sharing of climate, crop and economic models outputs. This application is useful to analyze different model outputs in various visualization formats and analysis results can be downloaded or shared and can be embedded within project websites in various formats. This is a one-stop solution for all The Agricultural Model Intercomparison and Improvement Project (AgMIP) model dataset storage, archiving, visualization and sharing. Different stakeholders like researchers, policy makers, students, universities can benefit from this tool. Desktop application can be distributed in CD/email and can be viewed on computers without internet connectivity.

Data repositories at ICRISAT

The ICRISAT data management team has come up with an innovative platform to support different data management needs of various program areas with limited resources in less time. Below are a few such data repositories for smooth functioning of the research data management. The repositories will ensure access of both institute programs and global scientific community to ICRISAT's research data to get successful results. This will showcase ICRISAT's confidence in open access data in research.

OAR and Dataverse

The data management team at ICRISAT has been working for increased usage of Dataverse platforms <http://dataverse.icrisat.org/dvn/> for data publications and data sharing to make research data openly accessible. Since its launch six Dataverse have been released with 54 studies and 443 files. This has been achieved through capacity building and training programs organized to educate scientists on the importance of open access to data and how to use these platforms. Based on the institute open access policy, ICRISAT set up an Open Access Repository <http://oar.icrisat.org > and a process of library-mediated archiving was established. The library consults the publishers' policy on each publication and accordingly uploads them in the repository. The library mediation is to ensure legitimacy in providing open access to ICRISAT's published research outputs. So far, the repository holds more than 6,700 research publications produced by ICRISAT researchers. The publications include journal articles, conference papers, book chapters, monographs and theses. In 2013, 312 research documents have been archived in the repository <http://oar.icrisat.org/view/year/2013.html>. The repository is institutionally defined. Hence, the publications that carry ICRISAT in the byline are archived in the repository. The Open Access Repository of ICRISAT abides by the CGIAR OA policy.

Village dynamics in south asia (VDSA) data warehouse

The longitudinal Village Level Studies (VDSA), of ICRISAT have for over three decades provided profound insights into the social and economic changes in the village and household economies in the semi-arid tropics of Asia and Africa. The VDSA project has recently developed VDSA Knowledge Bank with user-friendly data retrieval and on-line analytical processing features: The VDSA Knowledge Bank (technically a Data Warehouse) is developed. It is the first of its kind in the CGIAR system and also first in the world for management of rural household survey data. The Knowledge Bank uses Microsoft SQL Server and Microsoft Business Intelligence (MSBI) tools for its development. It has many useful and versatile features including easy and user-friendly data retrieval, and on-line analytical and processing (OLAP) features. It is the single source repository of all data including household survey data collected by ICRISAT since 1975 from six VLS villages in Telangana and Maharashtra, along with new data collected through the VDSA project from 42 villages in India and Bangladesh (from 2009 onwards). VDSA databases have been used by scholars in India and other countries for in-depth research in development economics, dynamics of rural economy and farming system. VDSA databases were frequently accessed by the global academic and research communities. Till July 2014, 653 unique users from 36 countries across Asia, Africa, Europe and North America have downloaded the newly released data. These include 306 students including 166 PhD students from 132 universities/ institutes around the world. Use of the VDSA datasets by Asian students and researchers has increased rapidly in recent

years. Currently, 311 researchers from India have been downloading the data on a regular basis.

ICRISAT- a where platform: Cloud based M&E and data sharing platform

This platform has been developed with a purpose to have an integrated data visualization and data sharing platform for Harnessing Opportunities for Productivity Enhancement (HOPE) Project and Tropical Legumes II (TL II) project. This application has weather modules which are very useful in comparison of the research analysis with climate impacts. This platform has three different types of data types including baseline, adoption and trial data. It supports the integration of TL II and HOPE project data for easy accessibility by project scientists. It also enables project scientists to provide controlled accessibility of their data to external users, including donors and other partners. In very less time the user can understand the research outcome and share the analysis with target community. Now, efforts are under way to enable French language support.

AGROBASE

This platform can be used as - complete data management side of breeding, including management of segregating material, for plant plans, analysis of data, reporting of data, and making field books. It has many of the desired capabilities including a good pedigree management system, generating experimental designs plan, generating and printing field layouts, importing data/trials from Excel and data loggers. This platform also has capabilities of doing basic statistical analysis by using GenStat, which is an add-on to get mean values with more sophisticated analysis. In addition, it can manage all trial data in underlying RDBMS with a query facility. This tool can be used to generate experimental designs for yield tests and for breeding nurseries. It has easy-to-use reporting tools for labels and field books.

Genetic resources

Currently, the Genebank Information Management System (GIMS) is a custom-built application that works well for the use of the genebank. The passport and characterization data on germplasm accessions is available on ICRISAT web and GENESYS. Grin Global is a suggested tool for genebanks. However, the implementation will require buy-in from the Genebank scientists and adequate training and support.

Integrated Breeding Platform (IBP)

The Integrated Breeding Platform (IBP) is a web-based, one-stop shop for information, analytical tools and related services to design and carry out integrated plant breeding projects with contributions from numerous partners and several key funders, coordinated by the Generation Challenge Programme of the CGIAR. The centerpiece of the IBP will be the Configurable Workflow System (CWS) – a fully integrated software system providing access to all the tools and data necessary to conduct day to day modern plant breeding. Access to tools and

data will be available via simple interactive graphical workflows that will be customizable for different crops and breeding strategies. ICRISAT is involved in developing the Molecular Breeding Design Tool for Marker assisted backcrossing projects, and the Genotyping Data Management System (GDMS).

EXPLOREit @ ICRISAT

EXPLOREit (exploreit.icrisat.org) is a new form of information management developed by ICRISAT gives the public a unique opportunity to explore and benefit from over 42 years of scientific data and information. EXPLOREit is now the main entry point to find any of ICRISAT's scientific information.

With a new revolutionary way of making information accessible called the multi-profiler concept provides an easier access to all of institute scientific research through multiple navigations and create easy-to-scan profiles on the subject areas.

EXPLOREit is not meant to be the source of all knowledge but brings all the scientific information ICRISAT has to offer freely and easily accessible to anyone!

Resource space

Much of the scientific information and data created by ICRISAT was not collated in any one area. To have it accessible through EXPLOREit, ICRISAT resource space team had to collect the information and data and find a suitable platform to keep it on as well as prepare it in a way that it would be able to automatically also feed into the appropriate sections of EXPLOREit.

Resource Space is now not only valuable for feeding material into EXPLOREit but also is a valuable internal resource that can be used to find required resources for reports and other requests. Resource Space provides a summary of all projects current & past and an inventory of data on success stories and videos for the public. Resource Space team is now in the process of adding posters, slideshows and photographs in this space.

Future action plans

The data management team will closely work with ICRISAT management and staff in further strengthening the data management capacities by bringing innovative approaches and new tools with partnership approach. The DMU will further improve the relations with the Consortium office and bring the technologies to ICRISAT. Learning is a continuous process and the DMU will train research staff on up-to-date management technologies and facilities. The data management team will work on the below recommendations for further strengthening research data management at ICRISAT.

Ethical committee to be established in all research programs and projects

Each research program and research project is recommended to have an ethical committee that is responsible for developing the research ethic guidelines and policies addressing sensitive data and data confidentiality, as well as

appropriate handling of research data. Projects should be reviewed based on their adherence to the accepted ethical standards of a genuine research study.

Clear guidelines on authorship attribution

To ensure that both technical and scientific staff are given the deserved credit for their work that has significantly contributed to the creation of the datasets they should be data authors. This includes all the people who played a key role in the following:

1. Conceiving and designing the field work in response to questions of recognized scientific importance and/or relevance for developmental impact and policy change.
2. Development and implementation of research designs, choice of methods, quality control on data collection.
3. Database design, data cleaning, validation and verification processes

The ICRISAT Data management team will guide the respective data authors in handling the program/project data management activities effectively.

Specific funds to publish legacy data

Legacy datasets that have high potential to contribute towards achieving the system level outcomes (SLO) should be archived and published. Most of these data sets are fully documented; however data formats and metadata need to be brought in line with today's requirements and standards. To ensure that these datasets can be made available to other ICRISAT scientists and partners, working in the CGIAR Research Programs, specific financial incentives need to be provided.

Changing institutional culture and effort recognition

Performance evaluations both at individual scientist level as well as project level need to shift from using simplistic indicators metrics such as numbers of papers, positions in lists of authors, and journals' impact factors towards assessing the quality of research itself. Research program directors need to put performance indicators in place that not only reward the excellent scientific writers the system has, but also the scientific and technical excellence that leads to the creation of the data, methods and ideas that are supposed to be communicated in the papers. Internal project reviews should take into account the technical rigor of the data collection procedures, the completeness of the data and its description, and alignment with existing community standards. Scientists and their field teams should be encouraged to produce peer-reviewed data papers.

ICRISAT research management has identified the importance of recognizing the efforts of scientists and research staff towards open data. Data management team is proposing a few methods of recognizing data authors - publish their name along with their contribution in the weekly in-house publication – 'ICRISAT Happenings'. This will also motivate scientists and research staff at ICRISAT to

works towards making research work open access for global community for future use.

Crop model data management platform

Cropping systems models require a minimum dataset to run the crop models and evaluate crop model simulation and outputs. Most of the models require data on soil, weather, phonological observations and crop management data. These datasets can often be used multiple times by different groups of modelers and experimentalists for different assessments serving various purposes. However, for various reasons, datasets are often underutilized because they are not properly stored or made available in required formats. Lack of data creates a bottleneck for many modelling efforts. Looking at these requirements, ICRISAT data management team is trying to make these important data available using open data repositories and also by developing a tool (online and offline application) for proper data visualization, sharing and to provide access to various clients for downloading valuable data on cropping system experiment and/or weather datasets in required formats for various locations in Asia and Africa.