IRRI

INTERNATIONAL RICE RESEARCH INSTITUTE

Good Practices for Managing Research Data

(IRRI Ref. No.: DPPC2007-66)

Proposal

submitted to the

Information and Communications Technology-Knowledge Management Initiative (ICT-KM)

May 2007

Contact:

Dr. Michael T. Jackson
Director for Program Planning and Communications (DPPC)
Telephone: +63 (2) 580-5600 ext. 2747 or 2513; Direct: +63 (2) 580-5621; Fax: +63 (2) 812-7689 or 580-5699
E-mail address: dppc-irri@cgiar.org Mailing address: DAPO Box 7777, Metro Manila, Philippines

---

Rice Science for a Better World

Improving CGIAR Effectiveness through Knowledge Sharing (KS)
A project of the ICT-KM Program of the CGIAR

Contact Information—Project Leader

Last name	Metz
First name	Thomas
Email address	t.metz@cgiar.org
Organization name	IRRI
Telephone/fax/Skype	Tel: +63(2)580-5600 ext. 2895 Fax: +63(2)580-5699, 891-1292 Skype: nurmetz

Project Details

Title of proposal	Good Practices for Managing Research Data
Area	Institutional KS
List of staff/partners and expertise working on the project	o Thomas Metz, international research fellow, research informatics (IRRI) o Emmali Manalo, officer, institutional information systems (IRRI) o NN, research informatics specialist (CIMMYT) o Contributions from specialists in data management, data analysis, software development, data encoding, and various research subject matter specialists
Project duration using grant funds	12 months
Country(ies) of Implementation	Philippines (IRRI), Mexico (CIMMYT)

PROBLEM DEFINITION

Most research conducted by CGIAR centers depends on the collection of primary research data. Effective and efficient research data management (RDM) requires special skills and experience, but there is frequently no systematic approach to teaching and training in good practices and skills for managing research data for new staff, or to update existing staff on new methodologies and technologies. Within an institute, these skills and experiences are scattered among staff with different specializations, for example, database managers, programmers, statisticians, data analysts, data managers, data encoders, principal investigators, and executing scientists. Good practices are often not recorded in the form of institutional guidelines and training materials of practical relevance for day-to-day research data management work or the induction of new staff members.

OBJECTIVES

- Develop, collect, record, and apply good practices in research data management at IRRI and CIMMYT
- Initiate and support communities of practice for research data managers at IRRI and CIMMYT

PROJECT IMPLEMENTATION

- The research informatics groups of IRRI and CIMMYT have been merged into the IRRI-CIMMYT Crop Research Informatics Laboratory (CRIL). This unit will provide the institutional framework for the execution of this project at both institutes.
- The central activity of the project will be the collaborative collection, development, and organization of content on good practices for managing research data, using wiki technology. Examples of wiki content on good practices for managing research data are
    - A frequently asked questions document
    - A guide to selecting the right tool for different problems
    - Disciplined use of MS Excel for data capture
    - Importing and managing research data in MS Access
    - Use of the R-statistics package for data visualization and analysis
- Didactically, the content of the RDM wiki will be structured as a cookbook. The recipes in this cookbook are presented as generic problems with detailed solutions that can easily be adapted to similar problems across different research domains. Such cookbooks have become very popular in the IT domain, as a complement to textbooks and reference books, as they respond to the frequent need to quickly find a solution to a specific problem.
- As a capacity-building method, the cookbook approach doesn't require research data managers to become experts in all relevant software packages or programming languages. With a basic knowledge and a set of relevant recipes, research data managers will be able to solve most of their research data management problems independently according to institutionally recommended good practices.
- The major part of the resources provided by this project will be used to build up the content of the wiki to a level where it can be used as a resource by research data managers in their daily work. Once this threshold of content has been reached, regular

users will be invited to become RDM wiki contributors and document problems and solutions that are specific to their research domains.

- A special effort will be made to collect good solutions to common problems from staff in the various disciplines involved in research data management and document them in the form of general but easy to understand recipes. Such a collection of recipes will define institutional good practices in research data management.
- Some staff at IRRI and CIMMYT have already used wiki technology as part of the Generation Challenge Program, the ICIS community, or the IRRI-CIMMYT Crop Research Informatics Laboratory.
- The RDM wiki will be visible and accessible only from within the IRRI and CIMMYT LAN domains. This restriction is deemed necessary in order to provide a protected incubator environment while content is initially built up and tested.
- A RDM wiki started at IRRI in March 2007, a few recipes were published, and feedback was sought. The response was very positive and it became clear that there is great demand for such a resource. The RDM wiki is not yet accessible from CIMMYT.
- The expertise to build such a resource is available at IRRI and CIMMYT, but the knowledge is widely distributed, generally not formally documented, and not presented in a generic problem-oriented way.
- The content of the RDM wiki will be published under an open-access license (Creative Commons) that allows free copying and the generation of derivative works.
- The project is planned for 12 months:
    - The first 9 months will be spent on building content on the RDM wiki and creating a user community that provides feedback on content, structure, didactics, and priority needs.
    - The last 3 months will be spent on building a community of practice (CoP) and RDM wiki contributor community.
    - Shortly after the initiation of the project, the project leader will visit CIMMYT to create awareness of the project and initiate and organize feedback and contributions from CIMMYT colleagues.

EXPECTED RESULTS

- RDM wiki accessible from IRRI and CIMMYT, containing a collection of problems and detailed solutions in research data management.
- RDM CoP at IRRI and CIMMYT that contributes to the RDM wiki and promotes its use within the institutes.
- Wide awareness and implementation of good practices in research data management at IRRI and CIMMYT by all staff involved.
- New research data management staff use the RDM wiki as the primary resource to learn institutional agreed-upon good practices.
- The RDM wiki will be maintained, updated, and enhanced as an institutional resource by the members of the RDM CoP beyond the CGIAR-funded project.
- As the content of the RDM wiki will be released under an open-content license permitting the creation of derivative works, the content can be reused in different forms, languages, media, etc., without requiring further explicit permission or negotiation.

INNOVATION

There is still little experience in using wiki technology within CGIAR centers to support communities of practice or institutional change processes. The openness and visibility of a wiki is often seen as a risk, rather than as an opportunity for increased participation and collaboration in communities of practice. IRRI has led the establishment and guided the use of several globally accessible wiki sites (e.g., GCP—http://cropwiki.irri.org/gcp, ICIS—http://cropwiki.irri.org/icis) for more than two years, and is now in a good position to use wiki technology internally for CoP support.

The problem-oriented cookbook approach has become very successful in commercial IT publications, but has not been systematically used as a capacity-building and pro-active support approach at IRRI or CIMMYT.

REPLICABILITY

- The use of wiki technology as a central resource for CoPs can be easily transferred to other centers and domains.
- The cookbook capacity-building and support approach can be transferred to different research domains, for example, good laboratory practices, good field experimentation practices, etc.
- The wiki software used in this project (MediaWiki) is freely available, easy to install and support, and widely known through its use for Wikipedia.
- As demonstrated by Wikipedia, the wiki technology and methodology are renowned for being very scaleable.
- The project can easily be scaled up by making it accessible from other centers' LAN domains or even globally.

TIMELINE

Month	1	2	3	4	5	6	7	8	9	10	11	12
RDM wiki content development												
RDM CoP initiation and support												
Travel to CIMMYT												