

Proposal for SP4 Commissioned Research

Task Title

Creation and maintenance of Generation CP Repository

Involved institutions

All Generation CP Consortium members
Agropolis, CIMMYT, IRRI, IITA, IPGRI/SGRP, ICRISAT, CIP

Proposed Task Leader

Raj Sood (IPGRI/SGRP)

Rationale/Justification

An enormous amount of data is being generated by the partners in the Generation Challenge Program. These data are stored and maintained at different locations, using different methods and standards. Most importantly, these data are not available for all partners to use. Acquiring, organizing and publishing the information on the Web through a Generation CP repository is critical for the successful completion of the tasks that require data from various sources. The repository will also play a crucial role in supporting the Generation CP network by deploying supporting infrastructure for a CP-specific MOBY¹ registry.

Furthermore, the repository will store and provide a collaborative environment for the development of templates, standards, data models, XML schemas, and evaluation reports of the analysis of software packages and tools.

Objective

The objectives of this project are to create a Generation CP repository for different datasets, data models, indexes, standards, XML schemas, evaluation reports from analysis of software packages, and to establish an online collaborative environment for the development of these components and their availability on the Web.

Methodology

The methodology proposed can be divided into four sequential steps: development of the Generation CP repository and collaborative environment, data collection and content management, implementation and deployment, and, maintenance modification and enhancement.

Phase-1: Development of Generation CP repository and collaborative environment

The first step of this phase will be to determine the user requirements for the Generation CP repository in collaboration with all the task and project leaders. This work will focus on meeting the storage requirements of data templates, models, XML schemas, standards and indexes. Furthermore, this repository will be used to meet the requirements of a CP-specific MOBY registry and for building a bridge between a MOBY registry and a UDDI² registry as used by GBIF³. Once

¹ BioMoby Project: <http://www.biomoby.org>

² Universal, Description, Discovery and Integration (UDDI)

³ Global Biodiversity Information Facility (GBIF)

requirements are defined, the appropriate repository will be created to accommodate the storage of different Generation CP data elements.

The second step of this phase will address design issues related to providing a collaborative environment for the development of data templates, data models, XML schemas and in making these elements available on the Web for CP partners. A Wiki type environment is proposed for this purpose. Through this environment the partners will be able to collaboratively develop the different CP related data components.

The third step of this phase will address design issues for creating a Web-based front end for data templates. The creation of tools transforming data models into templates will be undertaken in coordination with the other CP task and project leaders.

The Generation CP data repository and collaborative environment will be hosted and maintained by IPGRI staff.

Phase-2: Data collection and content management

A pro-active approach will be taken to collect data from CP partners and its integration into appropriate templates for storage. The development of the necessary tools to facilitate data migration from local systems to the templates will also take place in this phase.

The contents on the Wiki type environment will be updated regularly. This environment will provide yellow pages type information for all CP datasets on their availability and location and other related details. This will also apply to the electronic archiving of other CP related reports and documents.

The Wiki type environment will allow discussions related to the data models, XML schemas, and the standards to be facilitated by the moderators appointed for each category.

Dedicated staff and necessary time will be allocated to accomplish this task.

Phase-3: Implementation and deployment

Rapid implementation of the various components of the registry is necessary. Deployment of the Wiki type environment is a top priority so that the CP partners can commence developing XML schemas, templates and data models in a collaborative manner. This environment will facilitate community participation and will provide an easy to use interface. Two Wiki engines (Cowiki and Mediawiki) have already been tested and soon other partners will be invited to participate in the further testing.

Development of the front end for data templates will be done in coordination with CIMMYT. This front end will allow users easy access to data in the repository.

The CP specific MOBY registry will be deployed and hosted at IPGRI (Rome). Deployment of a dedicated server for hosting the Generation CP repository and MOBY registry is underway.

Phase-4: Maintenance, modifications and enhancements

The repository, collaborative environment and other components of the informatics platform will be maintained on a daily basis. Based on feed back from users and CP task leaders, necessary modifications and enhancements will be done on 'as

soon as possible' basis. Dedicated staff and the necessary time will be allocated to accomplish this task.

Activities

Phase-1: Development of Generation CP repository and collaborative environment

a. Requirements and design

The requirements of the Generation CP repository and the collaborative environment will be established in collaboration with the CP task leaders. The Generation CP repository and the collaborative environment will be designed based on these requirements. The requirements and design document will be circulated to a wider community to ensure that all user needs have been covered.

Phase-2: Data collection and content management

a. Data collection

The CP partners will be approached regularly for data collection. A 'help desk' service will be provided to the partners who need assistance to migrate their data from local databases to the appropriate data templates.

b. Content management

The contents of the Wiki type collaborative environment will be maintained regularly. The newly created XML schemas, standards, data model and software analysis reports will be published regularly. Agropolis, IITA, and CIP will evaluate the analysis software packages for SP1 and SP2. ICRISAT will evaluate the analysis software packages for SP3. These partners will engage in actual evaluation of the different software packages and will provide evaluation reports for publication on the Wiki collaborative environment.

Phase-3: Implementation and deployment

a. Wiki type environment

Two products (coWiki and MediaWiki) have already been tested. One of the two products will be implemented for further testing and validation by CP partners. The collaborative environment will provide discussion forums moderated by designated task leaders. The implementation of the Wiki environment will be done in close collaboration with Agropolis. Agropolis will contribute to the design and presentation of the contents on the Wiki collaborative environment. The Wiki environment will be implemented and deployed through a dedicated server at IPGRI (Rome).

b. Implementation of Generation CP repository for data templates

The Generation CP repository is foreseen as a database but other pragmatic approaches may also be adopted in storing the data templates. In coordination with the "Creation and maintenance of templates for Generation CP data storage in repositories" task lead by CIMMYT, one repository will be created to store all the CP data elements. An appropriate storage space in a secured server will be used. The data stored in the templates will be used to build indexes. The indexes will be used to create queries to search data in the Generation CP repository. The indexes will also be used by CP-specific UDDI registry.

c. Implementation of the repository for XML schemas, data models, and data standards

The repository of schemas, models and standards will be created. These elements will be published on a Wiki type environment and the CP partners will be invited to participate in the development of standards, schemas and the data models.

d. Implementation of web interface for data templates

An appropriate Web interface is required for the presentation of data templates. This interface will allow users to search and download data files. The work on this activity will be done in close collaboration with CIMMYT.

e. Implementation of a CP specific MOBY registry

The CP specific MOBY registry will be a component of the Generation CP repository. This registry will serve as a yellow pages for MOBY services for the Generation CP and will provide links with MOBY registry in Canada. Its development will be done in collaboration with MOBY experts and IRRI.

f. Implementation of the CP specific UDDI registry

The CP specific UDDI registry is a component of the Generation CP repository. The development of this registry will be done in collaboration with GBIF. This registry will be used for building a bridge between the Generation CP registry and the GBIF UDDI registry. This registry will provide an interface to publish passport and characterization data using UDDI technology. The entire GBIF community will be able to view and access passport and characterization data related to the CP. The necessary tools will be developed to bridge the gaps between MOBY and UDDI technology.

Phase-4: Maintenance, modifications and enhancements

The Generation CP repository and the Web based collaborative environment will be maintained and hosted at IPGRI. Dedicated staff members will do the necessary modifications and enhancements to the repository and collaborative environment regularly.

Dependencies and relationship with other tasks

This task is related to the following CP-SP4 tasks:

Tasks from 2004

1. Design central registry
2. Establishing fingerprinting and allele data repository
3. Developing priority software packages for mapping and QTL analysis

Tasks from commissioned research projects

1. Development of the Generation CP data models.
2. Application of MOBY for generation CP consortium.
3. Creation and maintenance of templates for Generation CP data storage in repositories.

Timeline and milestones

Activities	1 st Quarter	2 nd Quarter	3 rd Quarter	4 th Quarter
1. Requirements and design				
2.a Data collection				
2.b Content management				
3.a Wiki type environment				
3.b Implementation of Generation CP repository for data templates				
3.c Implementation of repository for XML schemas, data models, and data standards				
3.d Implementation of Web interface for data templates				
3.e Implementation of CP specific MOBY registry				
3.f Implementation of CP specific UDDI registry				
4. Maintenance, modifications and enhancements				

Project partners and their roles

The following list of partners is based on the tasks listed under SP4 Commissioned Research projects and the year 2004 task list. Partners that are not listed here will also be requested to contribute to this project.

Institute	Role
CIMMYT (Guy Davenport)	<ul style="list-style-type: none"> • Contribution in defining the requirements of repository and collaborative environment. • Contribution in designing the Generation CP repository. • Contribution in designing the front end for templates.
IRRI (Richard Bruskiewich)	<ul style="list-style-type: none"> • Providing documentation related to data models. • Providing documentation related to XML schemas. • Contribution in designing collaborative environment for developing schemas and models on line. • Contribution in designing central registry/index system for CP MOBY specific registry.
CIRAD (Brigitte Courtois)	<ul style="list-style-type: none"> • Providing software analysis reports and documents related to software packages for SP1 and SP2. • Contribution towards establishing Wiki type collaborative environment for developing schemas/models.
IITA (Francis Moonan)	<ul style="list-style-type: none"> • Providing software analysis reports and documents related to software packages for SP1 and SP2. • Contribution in defining requirements and designing the CP repository.
CIP (Simon Reinhard)	<ul style="list-style-type: none"> • Providing software analysis reports and documents related to software packages for SP1 and SP2. • Contribution in defining requirements and designing the CP repository.
IPGRI (Samy Gaiji)	<ul style="list-style-type: none"> • Providing information/documentation about XML schemas developed for passport and characterization data for publication. • Contribution in designing central registry for Web services. • Contribution in designing index system for Web services for MOBY and UDDI.
ICRISAT (S. Chandra)	<ul style="list-style-type: none"> • Providing software analysis reports and documents related to software packages for SP3.
Other Partners	<ul style="list-style-type: none"> • Making available Generation CP related data documents/reports for publication on Web.

Budget (US\$)

1. Requirements and design

	Project budget	In-kind contribution
Personnel		2,500
Supplies and Services		
Sub-total		
Sub Contract	6,822	
Overheads (3%)	178	
Grand Total	7,000	2500

2.a Data collection

	Project budget	In-kind contribution
Personnel	15,000	1,500
Supplies and Services	500	
Hardware	3,500	
Sub-total	19,000	
Overheads (18%)	3,420	
Grand Total	22,420	1,500

2.b Content management

	Project budget	In-kind contribution
Personnel	4,000	500
Supplies and Services	500	
Sub-total	4,500	
Overheads (18%)	810	
Sub contract	5,848	
Overheads (3%)	152	
Grand Total	11,310	500

3.a Wiki type environment

	Project budget	In-kind contribution
Personnel	1,000	2,000
Supplies and Services		
Sub-total	1,000	
Overheads (18%)	180	
Sub contract	5,848	
Overheads (3%)	152	
Grand Total	7,180	2,000

3.b Implementation of Generation CP repository for data templates

	Project budget	In-kind contribution
Personnel	8,500	2,500
Supplies and Services		
Sub-total	8,500	
Overheads (18%)	1,530	
Grand Total	10,030	2,500

3.c Implementation of repository for XML schemas, data models, and data standards

	Project budget	In-kind contribution
Personnel	4,000	1,500
Supplies and Services		
Sub-total	4,000	
Overheads (18%)	720	
Sub contract	4,873	
Overheads (3%)	127	
Grand Total	9,720	1,500

3.d Implementation of web interface for data templates

	Project budget	In-kind contribution
Personnel	5,000	2,000
Supplies and Services	500	
Sub-total	5,500	
Overheads (18%)	990	
Sub contract	4,873	
Overheads (3%)	127	
Grand Total	11,490	2,000

3.e Implementation of GCP specific MOBY registry

	Project budget	In-kind contribution
Personnel	3,500	750
Supplies and Services	500	
Sub-total	4000	
Overheads (18%)	720	
Sub contract	4,873	
Overheads (3%)	127	
Grand Total	9,720	750

3.f Implementation of GCP specific UDDI registry

	Project budget	In-kind contribution
Personnel	2,500	750
Supplies and Services	10,000	
Sub-total	12,500	
Overheads (18%)	2,250	
Grand Total	14,750	750

4. Maintenance, modifications and enhancements

	Project budget	In-kind contribution
Personnel	5,000	2,500
Supplies and Services	500	
Sub-total	5,500	
Overheads (18%)	990	
Sub contract	8,771	
Overheads (3%)	229	
Grand Total	15,490	2,500

5. Coordination

	Project budget	In-kind contribution
Personnel	22,500	
Supplies and Services	2,500	
<i>Sub-total</i>	25,000	
<i>Overheads (18%)</i>	4,500	
<i>Grand Total</i>	29,500	

Summary Budget

	Project budget	In-kind contribution
Personnel	71,000	16,500
Supplies and Services	15,000	
Hardware	3,500	
<i>Sub-total</i>	89,500	
<i>Overheads (18%)</i>	16,110	
Sub contracts	41,908	
<i>Overheads (3%)</i>	1,092	
<i>Grand Total</i>	148,610	16,500

	IPGRI/SGRP	CIMMYT⁴	AGROPOLIS	IRRI	IITA
Personnel	71,000				
Supplies and Services	15,000				
Hardware	3,500				
<i>Sub-total</i>	89,500				
<i>Overheads (3% and 18%)</i>	17,202				
<i>Grand Total</i>	106,702	11,695	11,696	9,746	8,771

⁴ The maximum overhead % agreed for the CP is at 18% total and 3% for pass through funds, i.e. the sub-contracts in this proposal. Sub-contractors will therefore include up to 15% overheads in their budgets.