



## Task 26: Development of GCP Central Repository

# GCP Central Repository

- Introduction
- Activities 2005
- Deviations from 2005 workplan
- Current status
- October-December 2005
- Proposal for Workplan 2006



# Introduction

## The GCP Central Repository

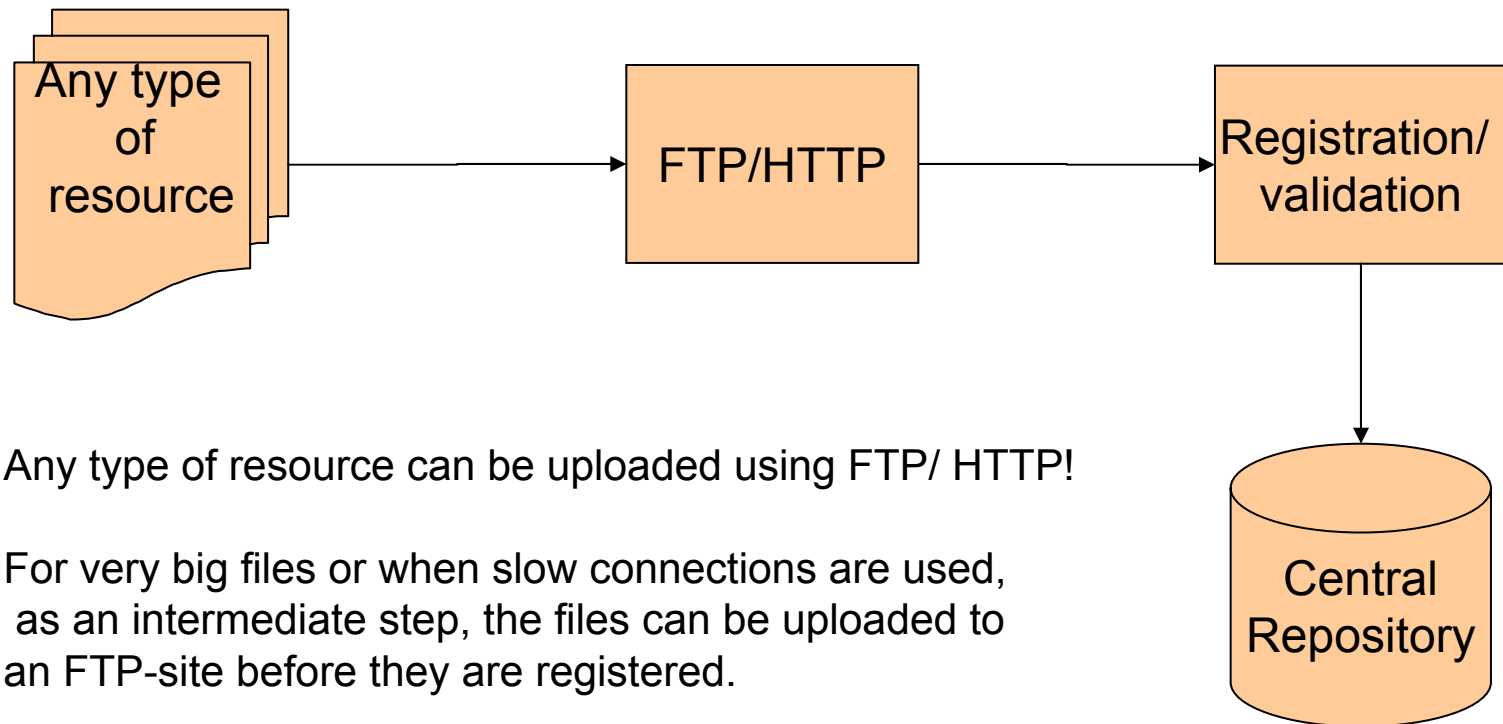
### The Central Repository provides:

- ✓ A directory ('yellow pages') where GCP resources can be registered and queried
- ✓ A facility to upload and download resources



# Introduction

## Uploading Resources through FTP/HTTP(1)



Any type of resource can be uploaded using FTP/ HTTP!

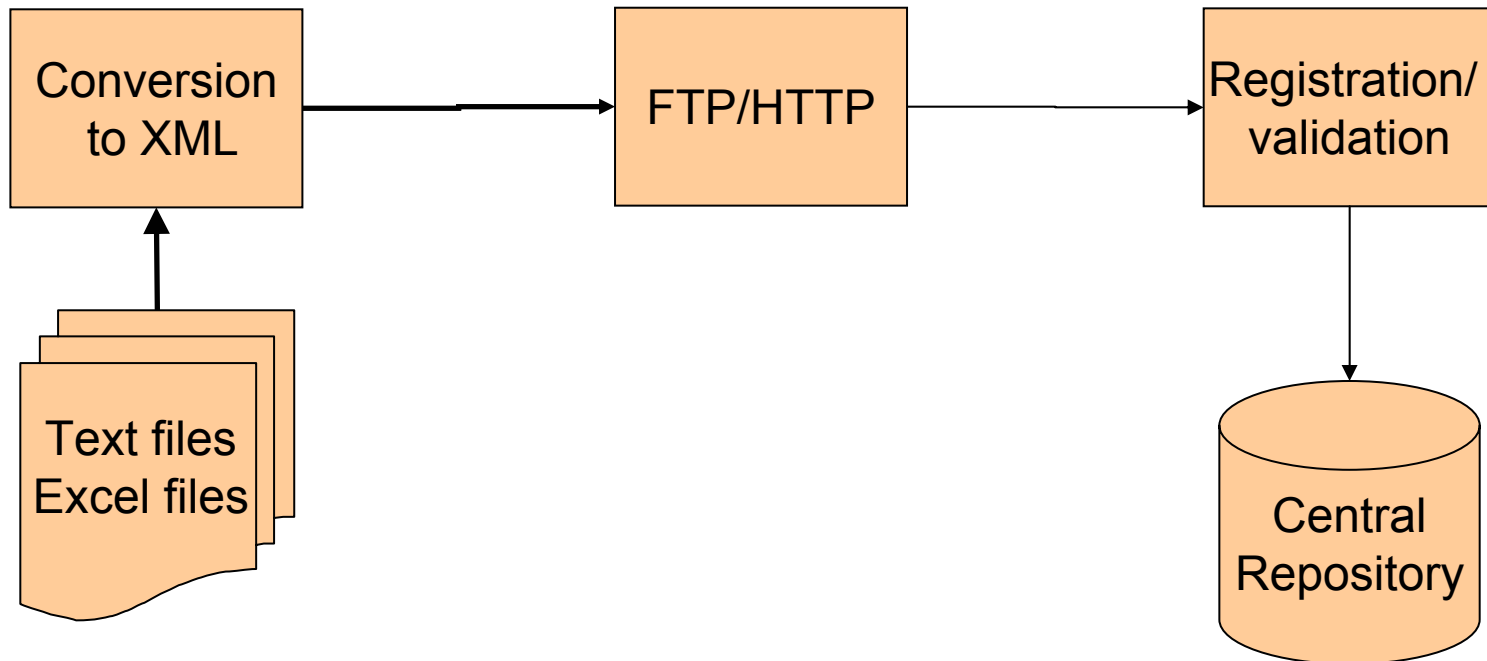
For very big files or when slow connections are used, as an intermediate step, the files can be uploaded to an FTP-site before they are registered.

FTP & Registration require user/passw login.



# Introduction

## Uploading Resources through FTP/HTTP(2)

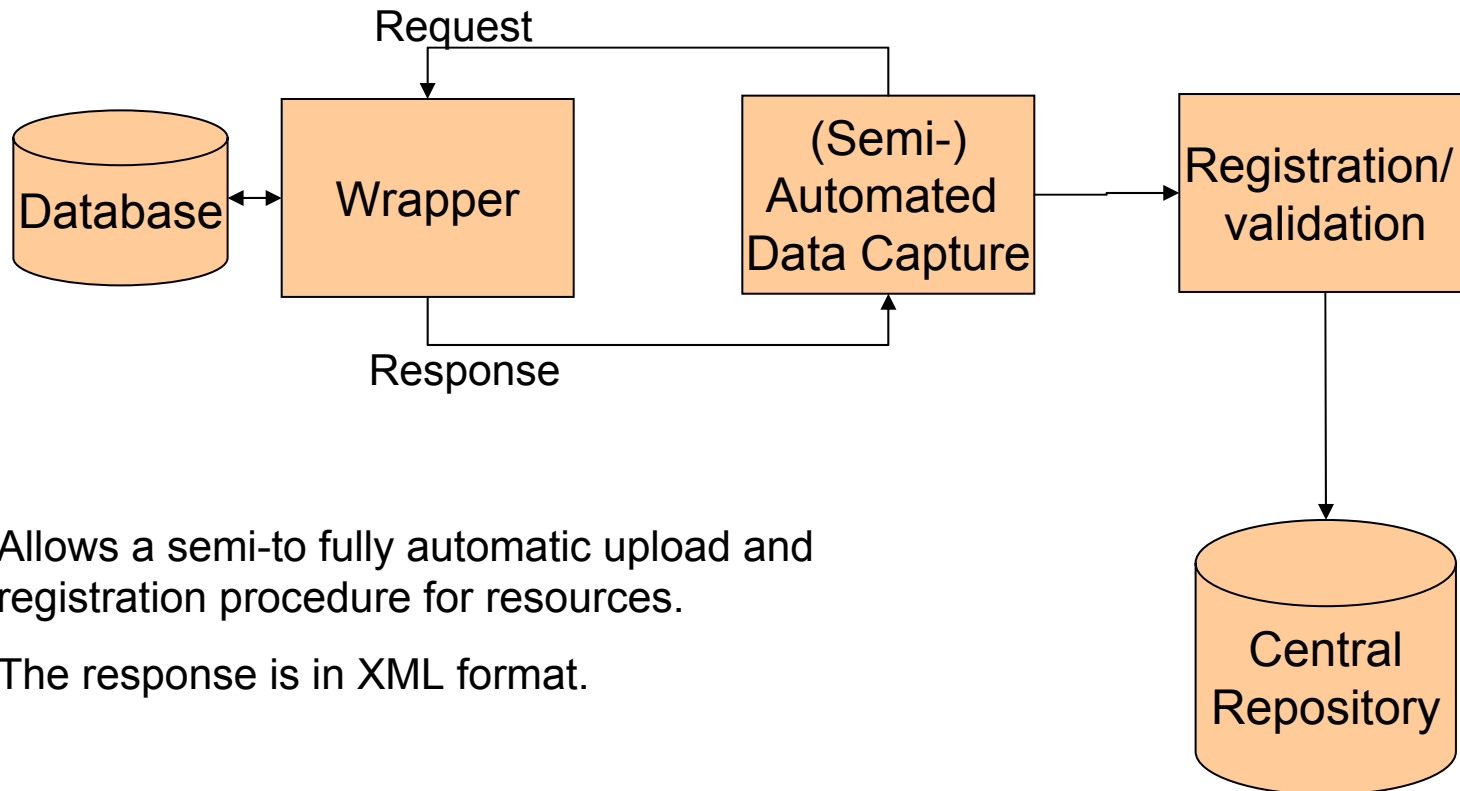


Templates: Conversion to XML format provides the data a well defined structure.



# Introduction

## Uploading Resources using web services



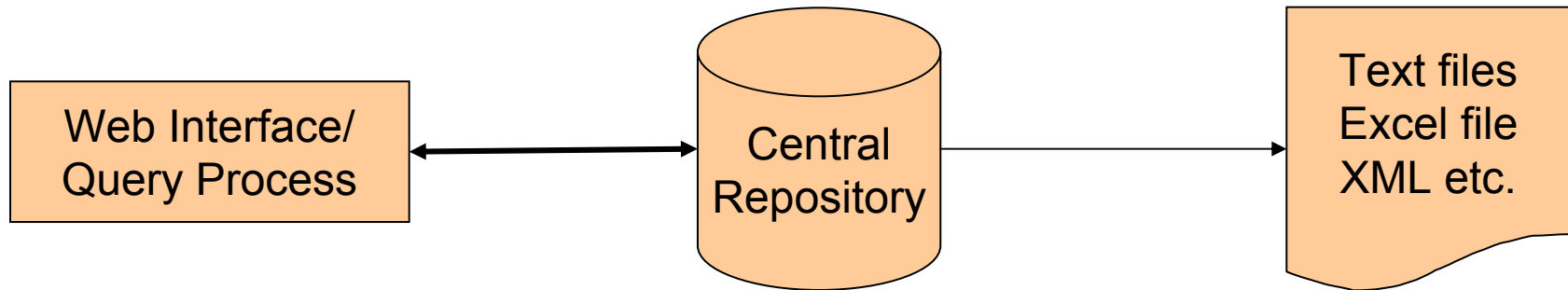
Allows a semi-to fully automatic upload and registration procedure for resources.

The response is in XML format.



# Introduction

## Central Repository-Query and Retrieval



### First phase:

- Queries on meta data
- Download of resources

**Second phase:** For resources which have a defined data structure (e.g. defined by XML schema), unit level data can be loaded.

To allow:

- Queries and visualization of unit level data (e.g. accessions)



# Activities 2005

Planned development and implementation in 2005

Phase	Q1	Q2	Q3	Q4
Requirements and design				
Data collection				
Implementation and deployment				
Maintenance				



# Activities 2005

Initial discussions on requirements and design (focused on requirements with regards the data templates)

Implementation of Central Repository highly dependent on other activities:

- Data collection: availability and use of data standards (templates) is essential. Templates developed during 2005.
- Choices for contributing technical solutions needed to be made (e.g. web services for automatic data harvesting).

First part of the year the development of the Central Repository had a slow start.

Efforts were redirected to further develop solutions on which outputs the Central Repository was dependent (e.g. testing BioCASE).



# Activities 2005

From May/June the pieces were coming together:

- Selection of BioCASE as distributed data delivery mechanism to e.g. GCP Central Repository
- Development of XML schema for gcp\_passport data for use with BioCASE (based on the passport template)
- Distribution BioCASE package to GCP Partners for installation
- First version repository submission data form developed for use with Templates (submission 'metadata')
- FTP site made available for submission of 'template' data
- Development GCP Central Repository Website



# Deviations from 2005 workplan

- Much of the development and implementation has been pushed back to second half of 2005
- Essential pieces (templates, BioCASE etc.) are NOW available to move faster on development and implementation
- Extra resources are committed to meet year-end deliverables



# Current status

## FTP/HTTP upload

- FTP site for intermediate storage of resources is available
- Parser to convert structured Excel/text files to XML is under development

## Automated datacapture

- BioCASE passport schema has been distributed to GCP partners for installation
- The Automated Datacapture feature under development

## Central Repository

- Functional version (i.e. upload, registration, query and download) of Central Repository website has been developed (<http://gcpcr.grinfo.net>)



# Current status

**Some screenshots.....**



## Upload file to the Central Registry

### Digital resource to upload

\* Select file to upload:

or select the file from your  
FTP folder:

### Resource provider

\* Provider institute:

\* Provider person:

\* Provider email:

\* Provided date: 2005-09-28  YYYY-MM-DD

### Resource source

\* Source institute: -- select or enter value below --

\* Source identifier:

\* Source person:

Source date:  YYYY-MM-DD

### Resource description

\* Title:

Description:

\* Type of resource:  other:

Gcp subprogram:

Project activity:

Crop:  other:

Status dataset:  other:

Digital format:  other:

Keywords:

Remarks:



### CENTRAL REPOSITORY

[Welcome to the Central Repository](#)

[Search files \(DEMO\)](#)

[Upload file \(DEMO\)](#)

[Search data \(NOT YET\)](#)

[Import data \(NOT YET\)](#)

[Central Repository Links](#)

[About the GCP-CR](#)

[LOGIN \(NOT YET\)](#)

### GENERAL LINKS

[Bioinformatics homepage](#)

[GCP Bioinformatics links](#)

## Search uploaded files

[\[Display\]](#)

### Resource provider

Provider institute:	<input type="text"/>	<input type="checkbox"/>
Provider person:	<input type="text"/>	<input type="checkbox"/>
Provider email:	<input type="text"/>	<input type="checkbox"/>
Provided date:	<input type="text" value="YYYY-MM-DD"/>	<input type="checkbox"/>

### Resource source

Source institute:	<input type="text"/>	<input type="checkbox"/>
Source institute:	<input type="text"/>	<input type="checkbox"/>
Source identifier:	<input type="text"/>	<input type="checkbox"/>
Source person:	<input type="text"/>	<input type="checkbox"/>
Source date:	<input type="text" value="YYYY-MM-DD"/>	<input type="checkbox"/>

### Resource description

Title:	<input type="text"/>	<input checked="" type="checkbox"/>		
Description:	<input type="text"/>	<input type="checkbox"/>		
Type of resource:	<input type="text"/>	<input type="checkbox"/>		
Gcp subprogram:	<input type="text"/>	<input type="checkbox"/>		
Project title:	<input type="text"/>	<input type="checkbox"/>		
Year:	<input type="text"/>	Type:	<input type="text"/>	<input type="checkbox"/>
Crop:	<input type="text"/>	<input checked="" type="checkbox"/>		
Status dataset:	<input type="text"/>	<input type="checkbox"/>		
Digital format:	<input type="text"/>	<input type="checkbox"/>		
Provided file name:	<input type="text"/>	<input checked="" type="checkbox"/>		
Keywords:	<input type="text"/>	<input type="checkbox"/>		
Remarks:	<input type="text"/>	<input type="checkbox"/>		

Check the boxes to the right for the columns you want to display in the results output.



### CENTRAL REPOSITORY

[Welcome to the Central Repository](#)
[Search files \(DEMO\)](#)
[Upload file \(DEMO\)](#)
[Search data \(NOT YET\)](#)
[Import data \(NOT YET\)](#)
[Central Repository Links](#)
[About the GCP-CR](#)
[LOGIN \(NOT YET\)](#)

### GENERAL LINKS

[Bioinformatics homepage](#)
[GCP Bioinformatics links](#)



## GCP - Bioinformatics - Central Repository

[\[Start\]](#)
[\[Search Files\]](#)
[\[Upload File\]](#)
[\[Search Data\]](#)
[\[Import Data\]](#)

### Search results (file details)

Your search criteria:

No search criteria given! All files displayed...

Title	Crop	File name	Details
	Test crop	<a href="#">README.TXT</a>	<a href="#">[Details]</a>
Central Repository Meta Data Descriptors	All	<a href="#">Central Repository Meta data descriptors.rtf</a>	<a href="#">[Details]</a>
Meta data descriptors		<a href="#">MetaData.doc</a>	<a href="#">[Details]</a>
FAO document with description of the GCP	All	<a href="#">GCP_mandate_crops.jpeg</a>	<a href="#">[Details]</a>
Common Schema	N/A	<a href="#">gcp_common.xsd</a>	<a href="#">[Details]</a>
2005 GCP Research projects		<a href="#">GCPProjects2005.txt</a>	<a href="#">[Details]</a>
Why MOBY	N/A	<a href="#">WhyMOBY.pdf</a>	<a href="#">[Details]</a>
2005 GCP research projects		<a href="#">GCPProjects2005.txt</a>	<a href="#">[Details]</a>
GCP web site banner	N/A	<a href="#">gcp-banner.jpg</a>	<a href="#">[Details]</a>
SP4 logotype		<a href="#">biq4.gif</a>	<a href="#">[Details]</a>
Finish flag bullet		<a href="#">Finland.gif</a>	<a href="#">[Details]</a>

### UNDER DEVELOPMENT (DEMO ONLY!)

The GCP Central Registry is still UNDER DEVELOPMENT!!

Feel free to browse the pages, but please do not be angry for the functions not yet developed or still unstable :-)

Please give us feedback on the design or planned features to [Dag Terje Endresen, d.terje@cgiar.org](mailto:Dag.Terje.Endresen@cgiar.org) and [Tom Hazekamp, t.hazekamp@cgiar.org](mailto:Tom.Hazekamp@cgiar.org)

# Current status

**Live Demo Friday Evening @ GCP ARM 2005**



# October-December 2005

## **FTP/HTTP upload**

- Parser to convert Excel/text files to XML developed
- Solicit/assist GCP partners to register resources

## **Automated datacapture**

- Automated data capturing feature for BioCASE installations operational

## **Central Repository**

- Repository activated and available on-line (End of October)



# Proposal Workplan 2006

## ***Objective***

- In 2006 the aim is to increase the depth and range of the resources it manages

## ***Strategy***

- Guidelines, manuals and a helpdesk function to assist in the submission of resources
- Pro-active approach to assist providers to register new resources
- Templates for new data types will be down streamed to providers as quickly as possible to allow the capturing of a broader set of data
- Development of new features (indexes and visualization tools)



# Proposal Workplan 2006

## Work packages:

- 1) Technical maintenance and management of the GCP Repository***
- 2) Functional development of GCP Repository***
- 3) Building up the Central Repository's resource collection***
- 4) Content management***
- 5) Helpdesk, manuals and guidelines***



# Proposal Workplan 2006

## Work package 1

### *Technical maintenance and management of the GCP Repository*

#### Outputs

- Documented management procedures for the Repository
- Proper management of Central Repository

#### Implementation

The server on which the Central Repository runs is located at IPGRI. It is proposed that the technical maintenance and management of the Central Repository is therefore performed by IPGRI.



# Proposal Workplan 2006

## Work package 2

### *Functional development of GCP Repository*

#### **Outputs**

- Indexing system for unit level data maintained by GCP Repository
- Implementation of visualization tools to assist in the discovery of unit-level data

#### **Implementation**

Technical implementation by IPGRI. To determine the core data types to build appropriate indexes, inputs will be sought from the lead institutes that develop templates for the various data types (Agropolis, CIMMYT, IITA, IRRI and SCRI).

For the visualization tools it is proposed that CIP and IPGRI work closely together. CIP has demonstrated experience with this type of tools.



# Proposal Workplan 2006

## Work package 3

### *Building up the Central Repository's resource collection*

- 1) Pro-active approach to providers to stimulate submission of resources**
- 2) Downstream new standards developed within GCP**

### **Outputs**

- Increased number of resources submitted to the Central Repository
- New standards down streamed to potential providers resulting in a broader set of data types captured in the Central Repository

### **Implementation**

It is proposed that IRRI, CIP and IPGRI work together on this task for their specific time-zone region. Coordination with CIMMYT on release of new templates.



# Proposal Workplan 2006

## Work package 4

### *Content management*

#### **Outputs**

- Controlled vocabularies for types of meta, registration and unit level data
- Enhanced validation rules for registration data and unit level data

#### **Implementation**

- It is proposed that IPGRI leads the content management as it hosts the Central Repository databases. In addition inputs will be sought from the lead institutes that develop templates for the various data types (Agropolis, CIMMYT, IITA, IRRI and SCRI) to develop and or implement the required controlled vocabularies and validation rules.



# Proposal Workplan 2006

## Work package 5

### *Helpdesk, manuals and guidelines*

#### Outputs

- Documented guidelines for data submission and use of the Central Repository
- Helpdesk established
- A limited amount of on-site visits is foreseen to provide hands-on support in situations where on-line assistance is not sufficient

#### Implementation

It is proposed that IRRI, CIP and IPGRI work together on this activity. IPGRI's role will be to develop the appropriate documentation and provide helpdesk support in their time-zone region. IRRI and CIP would be in a good position to provide help desk support to the GCP partners they have trained in installing the BioCASE installations in their respective time-zone region.



# Proposal Workplan 2006

## **Work packages & linkages to specific SP 4 projects:**

### ***1) Technical maintenance and management of the GCP Repository***

### ***2) Functional development of GCP Repository***

With regards to creation of appropriate indexes:

- Creation and maintenance of templates for the GCP data storage in repositories (Davenport)

### ***3) Building up the Central Repository's resource collection***

With regard down streaming new schemas:

- Implementation of Web services technology in GCP (Skofic)
- Development of GCP domain models (Bruskiewich)
- Creation and maintenance of templates for the GCP data storage in repositories (Davenport)

### ***4) Content management***

With regard to application of quality of content:

- Improvement of quality of existing databases (McLaren)

### ***5) Helpdesk, manuals and guidelines***





End