



# CIP – RIU selected activities

Information systems for crop and  
genebank data management

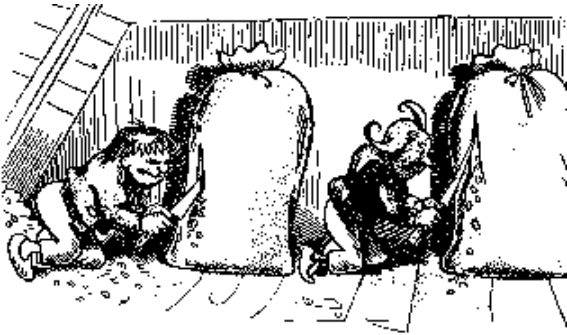
Edwin Rojas (Software Developer Chief),

Reinhard Simon (RIU Head)

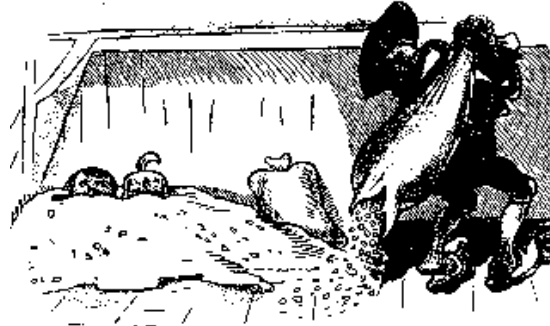
ICIS workshop 2006, CIMMYT

# The Research Informatics story

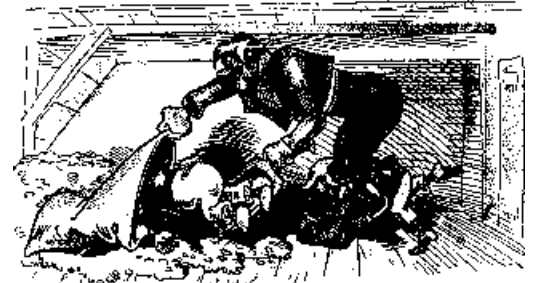
## Research



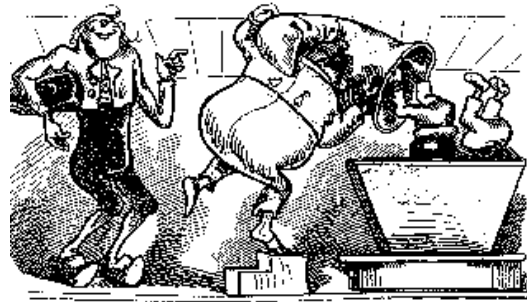
**Problematic situation**



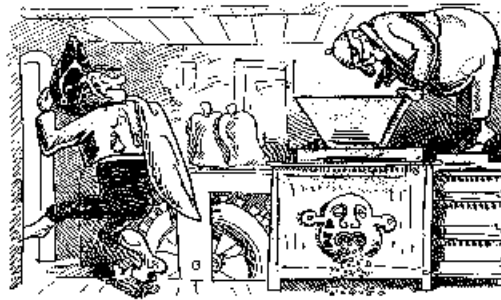
**Hypothesis**



**Data capture**



**Data/information management**



**Analytical processing**



**Standardized reports**

## Research informatics

Presentation to CIP annual meeting 2004  
Reinhard Simon



**Publication for diverse uses/uptake**

Adopted from: W. Busch:  
Max und Moritz, 1865

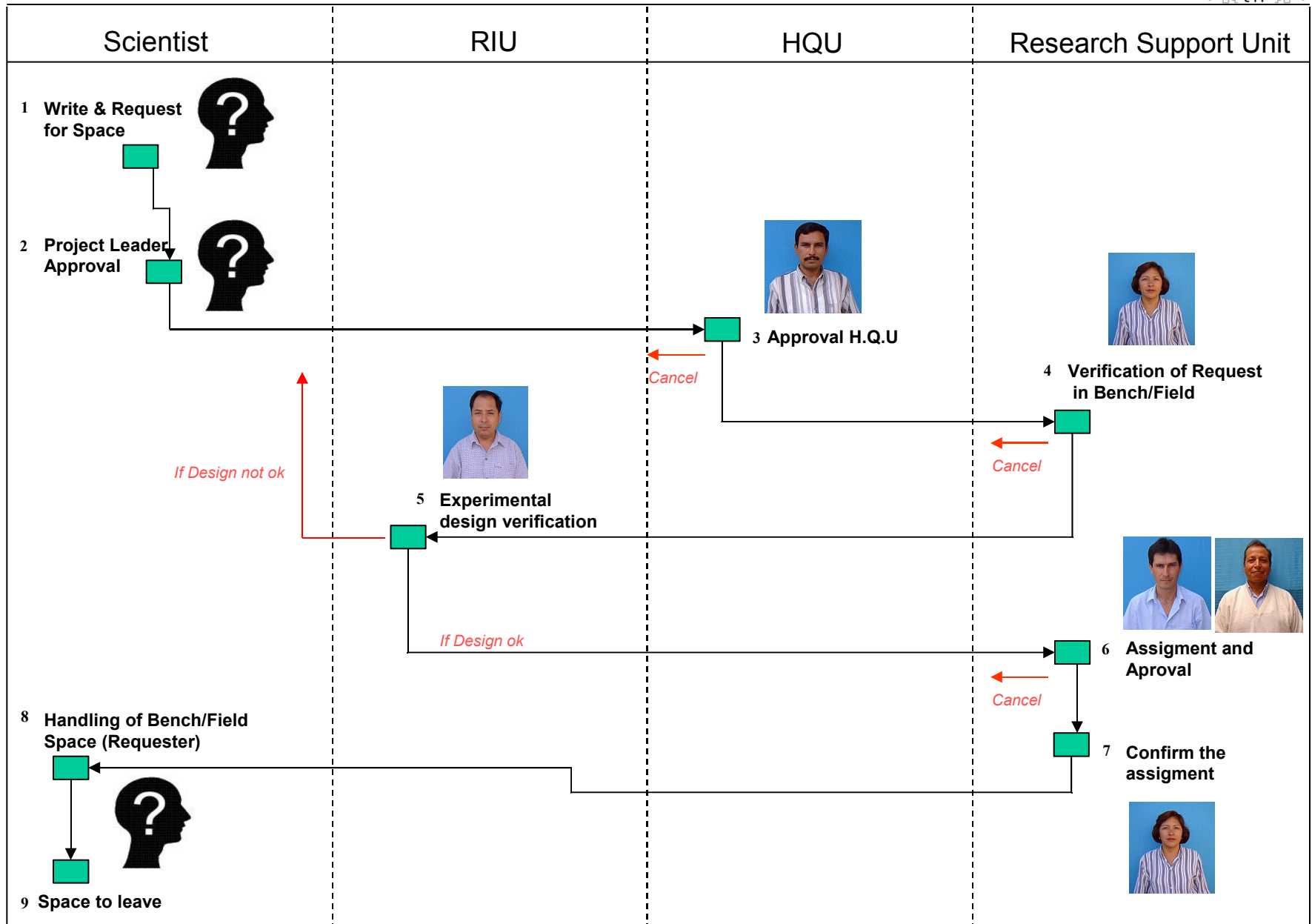
# Selected RIU Activities

- Workflows for:
  - CIPSTATIONS (Request Space for Field & Greenhouse)
  - CIPGADC (Request Germplasm Distribution, Acquisition & Cleaning),
  - CIPVIR (Request Pathogen Diagnostic: Virus & Viroids)
- CIPPEX (Register Experiments) [research.cip.cgiar.org/cippex](http://research.cip.cgiar.org/cippex)
- CIPSTAT (Analysis Experiments) [research.cip.cgiar.org/cipstat](http://research.cip.cgiar.org/cipstat)
- LIMS for molecular marker lab and “quality” lab
- CIPTCL (In-Vitro Genebank Management)
- DIVA-GIS (Free GIS tool) [www.diva-gis.com](http://www.diva-gis.com)
- Data warehouse for quick & flexibility access to explore data and quality control
- Software development and collaboration tools
- Outlook – harmonization with ICIS database schemes and tools

# Benefits of Workflow System

- The power and potential of Workflow
  - As a tracking activities system
  - Reduction of papers, telephones and mails
  - End users can visually orient themselves on the progress of their work
  - Improve archiving and auditing (time activities & cost statistics)
  - If research rules changes then the workflow engine can be updated with the new Workflow by final user
  - Granularity security for each activity in the Workflow

# Workflow Process: Request Space Field, Greenhouse



# Workflow Front-End: CIPSTATIONS



## Experimental Station Management System (CIP-STATIONS)

Maintenance ▶ Register a Request ▶ Search & Reports ▶ Help ▶



Experimental Station Management System

### Request for Field Space

Current Activity: Request Finalized

Have files: Yes

Year Number Next Activity

Year of request:  Number of request:

Name Leader:  Subleader:

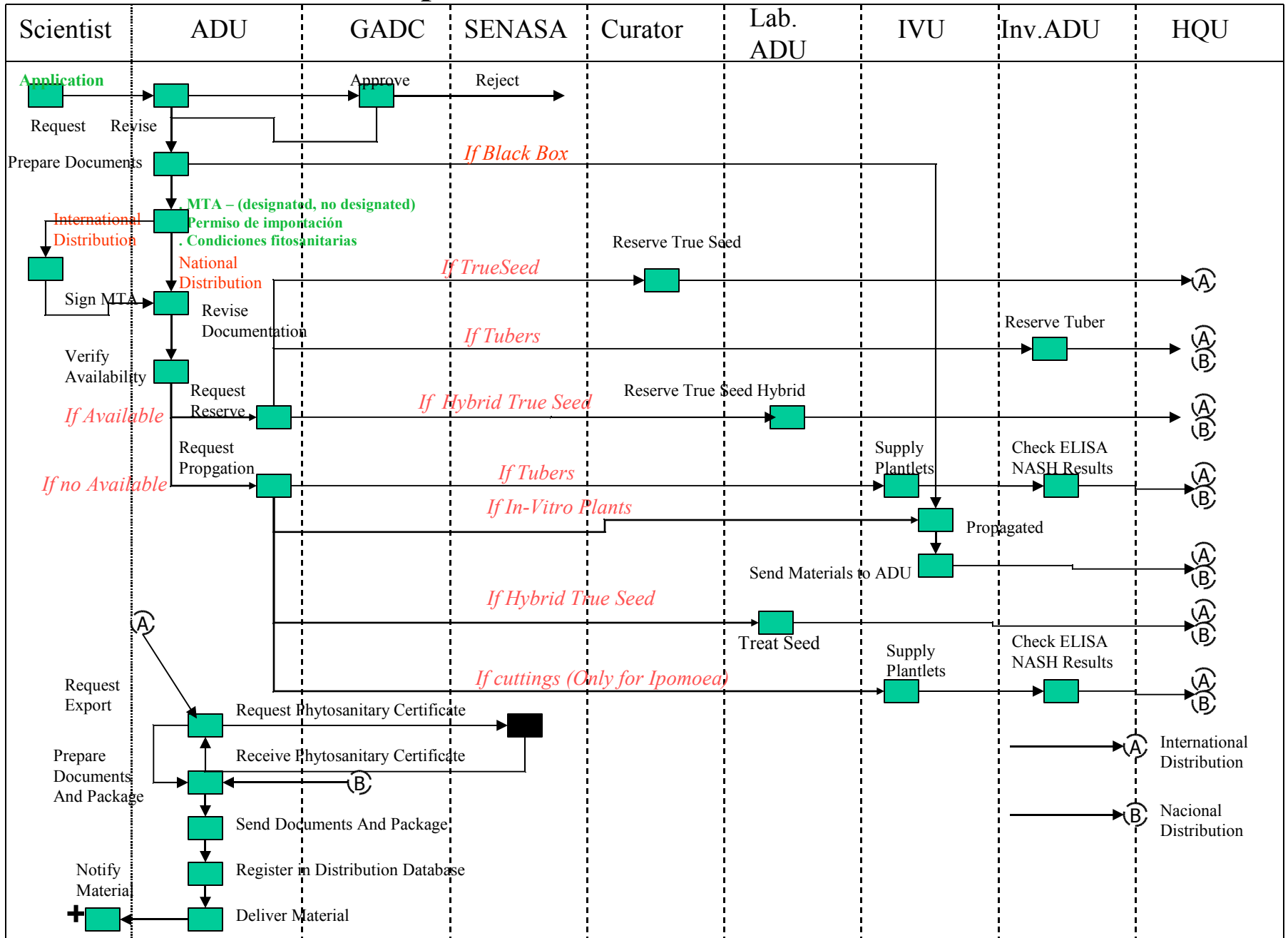
Name responsible approval request:  Date request:  mm/dd/yyyy

Name Assistant:  Project:

Crop:  Other crop:  Type Use:

Experimental Stations:

# Workflow Process: Germplasm Distribution

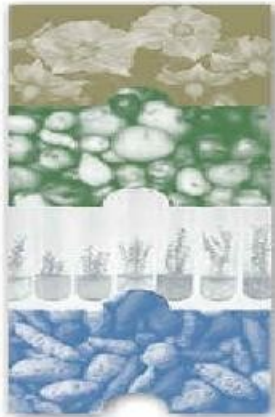


# Workflow Front-End: CIPGADC

Germplasm Acquisition, Distribution, Conservation and Pathogen Testing System (CIPGADC)

Maintenance ▶ Register a Request ▶ Search & Reports ▶ Utils ▶ Help ▶

## Welcome!



CIPGADC is an informatic tool for managing the Acquisition, Quarantine, Testing, Cleaning, Maintenance, Multiplication, and Distribution of Germplasm in CIP.

Germplasm data provided by 3 Service Units ([ADU](#), [IVU](#) and [PHOU](#)) are integrated into a system that allows users to place their germplasm request online. Follow up status of material acquired or placed under quarantine; pathogen tested; cleaned up; maintained as invitro or search capability and can p

### Search a distribution of genetic material

CIPGADC evolved from the and Distribution Committee standards for handling plant regulations concerning plant

We welcome any comment

2006

All  Year of distribution : All

All  Destination: All

All  Status: All

All  Scientist : All

All  Institution: All

---

n	Date of ADU Request	Date Expected Delivered	Date Distribution	Crop	Consignee	Institution	Country	Scientist	Status
	01/05/2006	02/28/2006	03/06/2006	Potato	Rosario Falcon	CIP - Lima	Peru	Enrique Chujoy	Distributed
	01/05/2006	03/15/2006	04/11/2006	Potato	Rosario Falcon	CIP - Lima	Peru	Enrique Chujoy	Distributed



# CIPPEX [research.cip.cgiar.org/cippex](http://research.cip.cgiar.org/cippex)

- Based on tool for project management (PHProjekt)
- Phenotype data management inspired by ICIS
- Genotype data management inspired by GERMINATE
- Web Public Flash Videos for Training

CIP databases:  
passport, charact, eval.  
data

Project and experiment  
management:  
PHProjekt

Breeding data:  
adaptations from  
ICIS

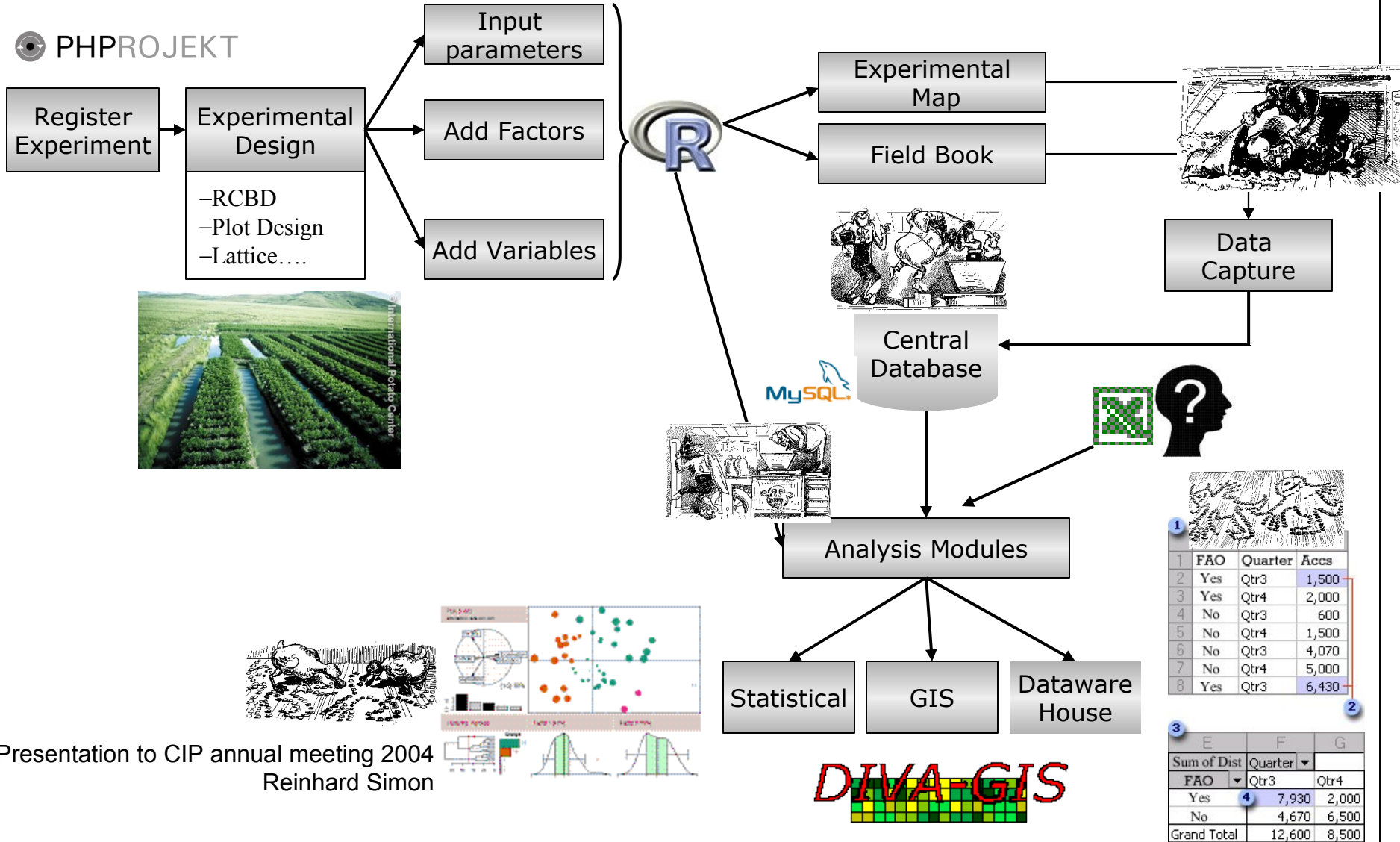
CIP-Workflow  
Engine

Resource scheduler:  
MRBS

Molecular data:  
adaptations from  
Germinate

# Diagram of the Experimental Design & Analysis Generator Module for the CIP-PEX System

## PHPROJEKT



Presentation to CIP annual meeting 2004  
Reinhard Simon































**DIVA-GIS**

# CIP Mobile/Barcode Infrastructure Status 2006-01

 Barcode

Pockets  

LAN Wireless 

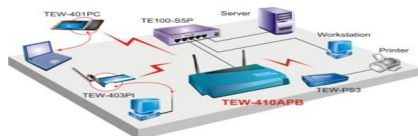
Areas	Potato	Sweet potato	ARTC
<b>LIMA</b> 	In vitro Chamber	  	  
	Cold Chamber for Seeds	Pending 	Pending 
	Cold Chamber for Tubers	Pending 	—
	Greenhouses	 	 
	Field Laboratories (Molecular/Processing)	Pending 	Pending 
<b>JUNIN</b> 	In vitro Chamber	Pending 	 Pending 
	Greenhouses	 Pending 	Pending 
	Field	 Pending 	Design completed by ITL 
	Storage Room	Pending 	—
	Laboratories	 Pending 	Pending 

Notes: The barcode includes the use of labels, thermal printers and reader guns. The symbol “—” means that the specific crop does not are the respective facility.



First Pilot in In vitro Potato

2000



Infrastructure Design for all In vitro & Greenhouse

2003 January



Infrastructure for all In vitro & Lima-Greenhouse

2003 November



Applications

2004

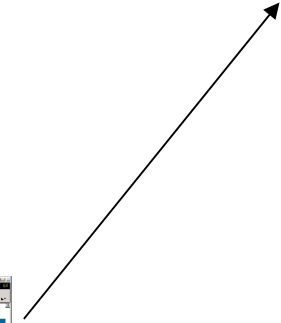
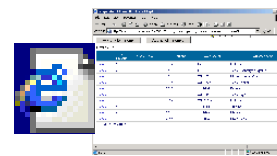
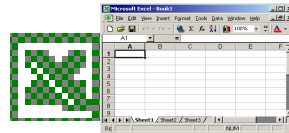
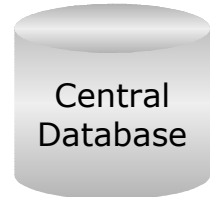
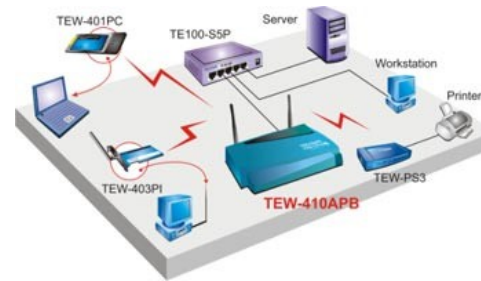


Improve Infrastructure and Wireless Covertures (from 11MB to 54 MB)

2005-2006



# Environment Experiment Data Capture with Mobile Device



# In-Vitro Genebank Performance Indicators

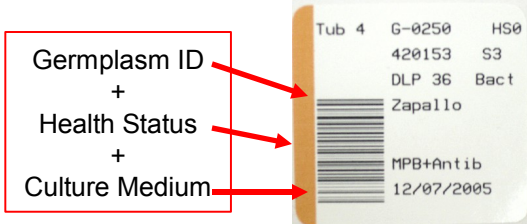
In-Vitro Genebank Activities	Barcode and Mobile Solution Technologies	Prior to Implementation Technologies
Locate a material by any staff	5 seconds	1800 seconds
Create a list for a “grid” of 15 entries	30 seconds	90 seconds
Print 90 labels for a “grid”	20 seconds	120 seconds
Clients request germplasm availability for distribution	10 seconds (by on-line web)	1 or 2 days (by email, phone)
Inventory report for stock and locations	1.5 weeks by 2 staff	6 weeks by 2 staff
Migration to barcode	100% of germplasm use autoadhesive labeled with barcode identification	100% of germplasm used paper labeled and eye identification

# Genebank Innovations for Better Performance

## 1. Mobile Computers (Pocket PCs)



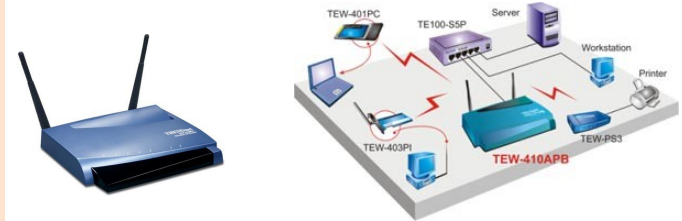
## 2. Barcode (1D)



## 3. Thermal Printers



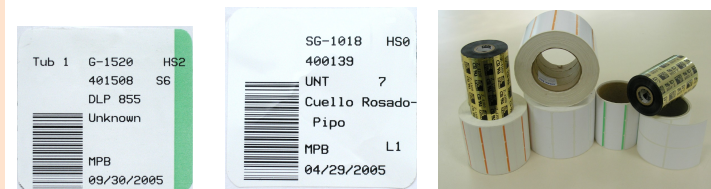
## 4. Wireless Networking



## 5. Hand Barcode Readers



## 6. Media: Labels and Ribbons



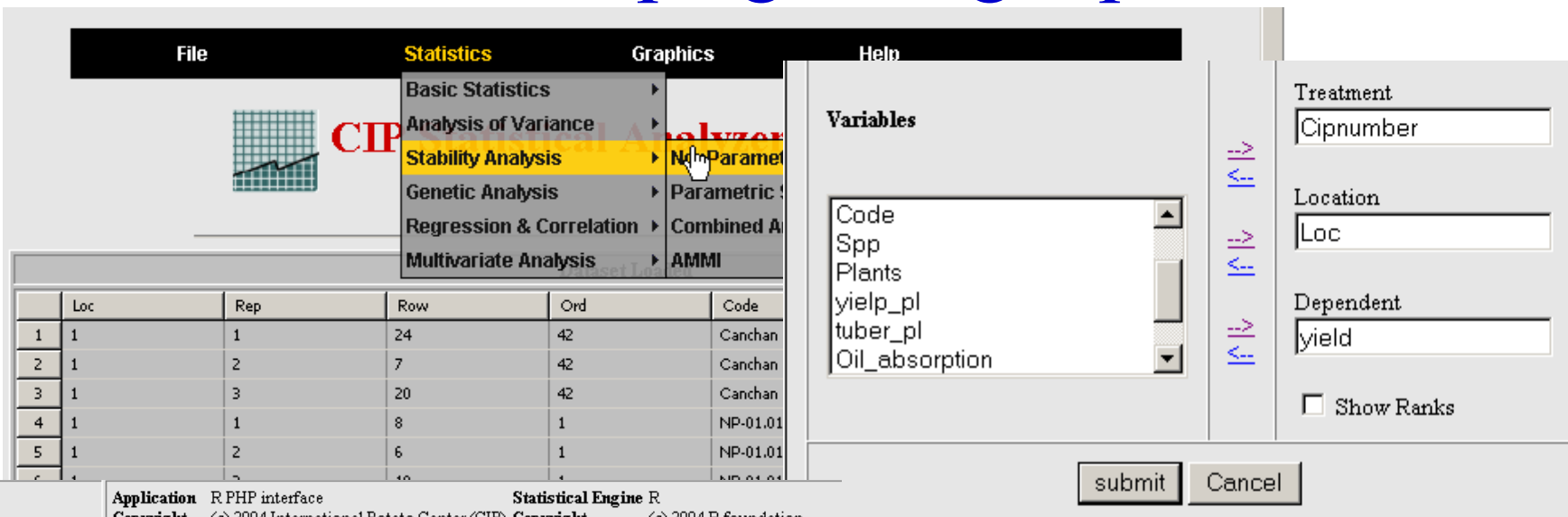
# CIPSTAT

[research.cip.cgiar.org/cipstat](http://research.cip.cgiar.org/cipstat)

- Web interface for users (via web forms) and software (via WSDL web services) to statistical routines
- Statistical routines packaged in a library 'Agricolae' written in R
- Presently access to 29 routines covering frequently used at CIP for experimental planning and analysis
- Web Public Flash videos for Training

# CIPSTAT – screenshot

[research.cip.cgiar.org/cipstat](http://research.cip.cgiar.org/cipstat)



**File**    **Statistics**    **Graphics**    **Help**

Basic Statistics  
 Analysis of Variance  
 Stability Analysis  
 Genetic Analysis  
 Regression & Correlation  
 Multivariate Analysis

Variables

Treatment: Cipnumber  
 Location: Loc  
 Dependent: yield  
 Show Ranks

submit    Cancel

	Loc	Rep	Row	Ord	Code
1	1	1	24	42	Canchan
2	1	2	7	42	Canchan
3	1	3	20	42	Canchan
4	1	1	8	1	NP-01.01
5	1	2	6	1	NP-01.01

Application R PHP interface      Statistical Engine R  
 Copyright (c) 2004 International Potato Center (CIP)      Copyright (c) 2004 R foundation  
 Developed by CIP-RIU  
 Contact [cip-riu@cgiar.org](mailto:cip-riu@cgiar.org)      Contact [www.r-project.org](http://www.r-project.org)  
 Version 0.9.0      Version 1.8.1

[View Textual Report](#)

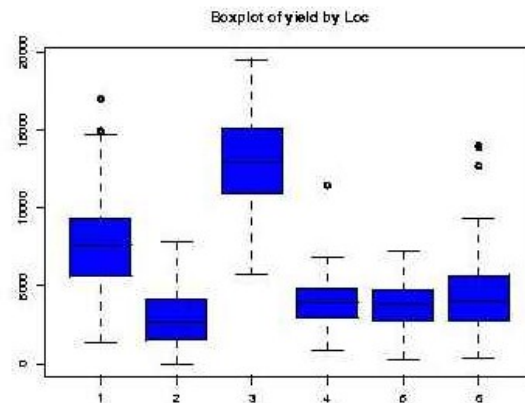
[View Graphics](#)

[Download](#)

Cipnumber x Loc Table of yield means computed from data

	1	2	3	4	5	6
374080.5	8966.667	3382.8333	15450.373	4300.000	4150.0000	NA
380389.1	7283.333	2767.5000	15058.053	2133.333	2183.3333	7716.667
700234	9200.000	3479.5000	15083.170	3116.667	4833.3333	7183.333
700313	6083.333	4897.5000	12408.667	4858.333	3616.6667	3966.667
700787	10833.333	2709.5000	16299.170	5725.000	3833.3333	6016.667
701165	7366.667	1194.1667	8335.000	1675.000	3250.0000	1933.333
701273	9916.667	4170.6667	15804.667	5833.333	5166.6667	6183.333
701515	11183.333	5517.5000	14497.667	6025.000	4016.6667	5133.333
701675	8050.000	3639.0000	17256.080	3833.333	5100.0000	3266.667
701997	10366.667	2194.1667	12865.980	4125.000	5186.6667	6233.333
702363	8216.667	6151.5000	16683.437	4491.667	5350.0000	4383.333
702395	7283.333	3682.0000	13662.067	3275.000	3700.0000	3116.667

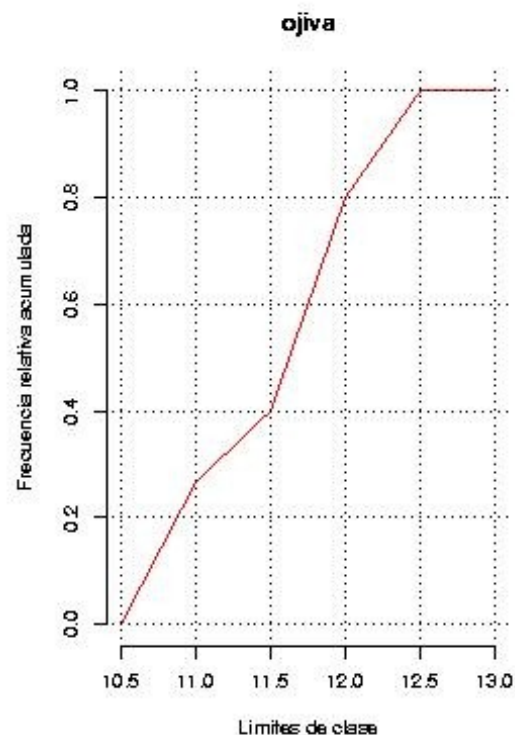
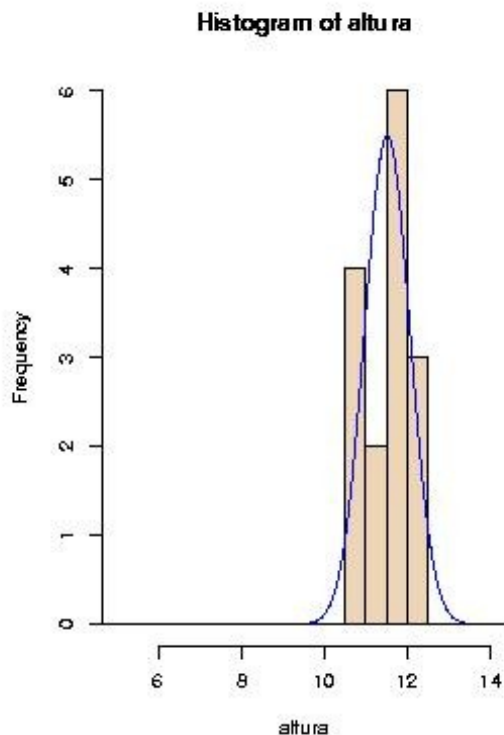
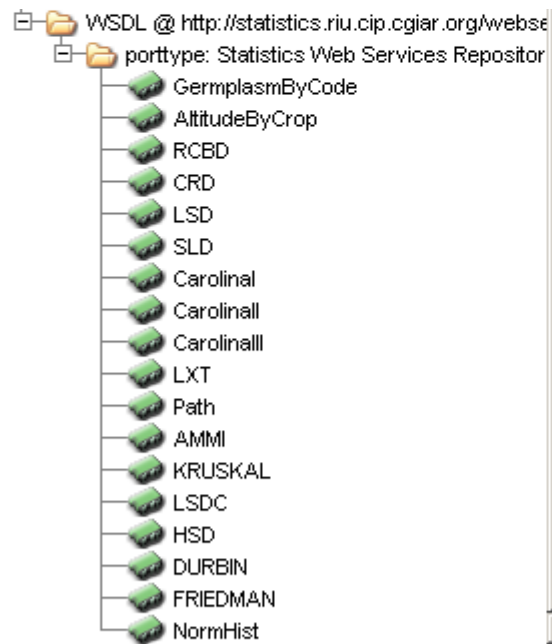
[Download this image in a Postscript Format](#)



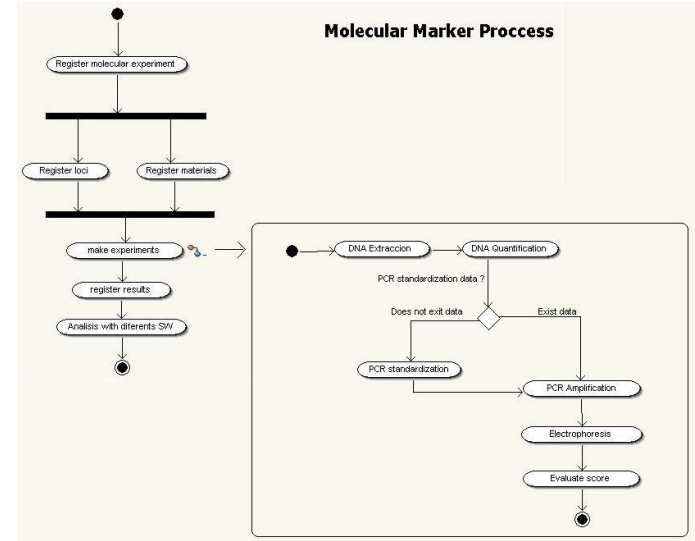
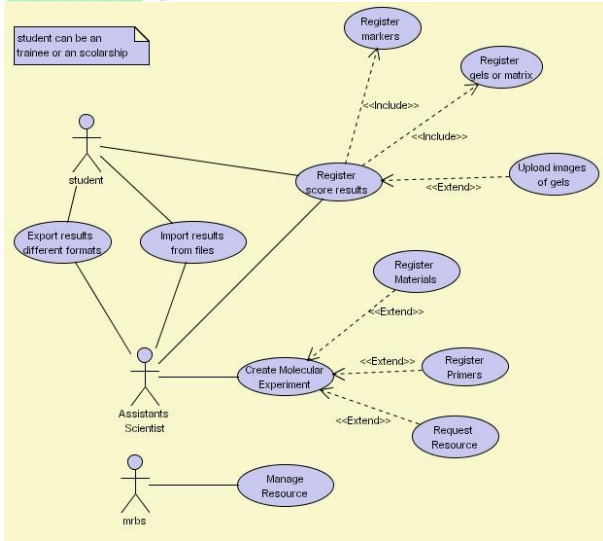



# CIPSTAT – screenshot

Interface for tools via WSDL – e.g. Taverna



# CIPPEX – LIMS molecular data





Short Name:

Begin:  End:

Crop:  Category:

Leader:

Student:

Type of experiment:

Type of marker:

Status [%]:

Priority:

Sub-Project of:

List:

Dependency:


**Select Primer**

**Choose the materials you require**

**Select matrix**

**Select gel**

IB-R08	401396	401398	401391	401392	400098	401393	401498	401499	401447	401522	401151	401152	401150	401153	401154	401223	400397	401224	401225	401228	401227	401228	401132	401076	401079
200	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
207	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
214	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0
216	1	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
217	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
218	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
220	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
224	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
226	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0



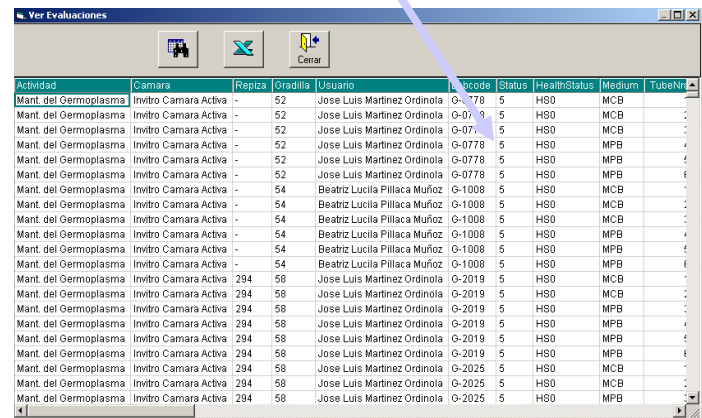
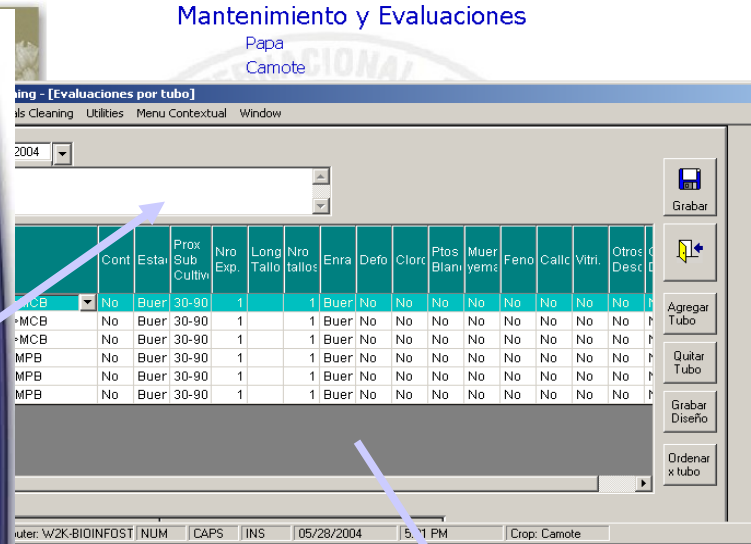
# CIPTCL – LIMS for GR

Sistema de Administración de Recursos Genéticos

## Sistema de Administración de Recursos Genéticos

### Mantenimiento y Evaluaciones

Papa  
Camote

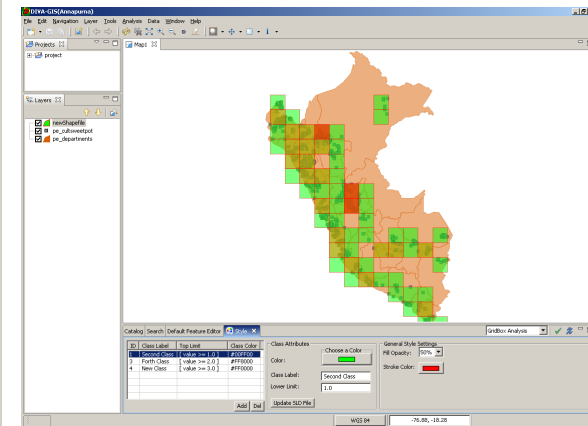
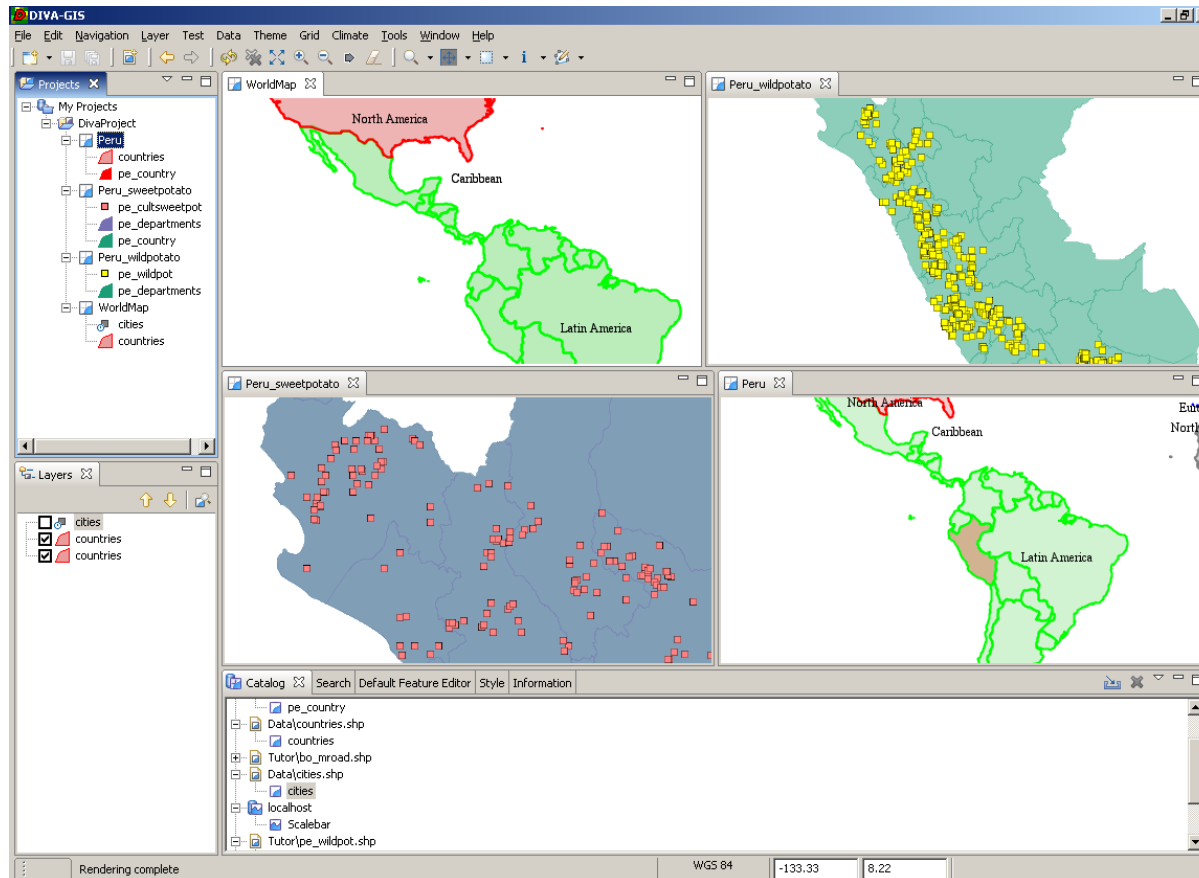


# DIVA-GIS overview

[www.diva-gis.com](http://www.diva-gis.com)



DIVA-GIS is a free and Open Source geographic information system (GIS) application that is used specially on GIS support for genebank curators & breeders.



# Data warehouse

- MS based solution for genebank and GADC
- OS based solution for breeding experiments (and others)

# Data warehouse – CIP Intranet



This page and the dynamic

## Dynamic Reports for

With **Dynamic Reports for Genetic** data, compare information and view reports from Microsoft Excel. The reports show CIP's genebanks since 1979 and Genetic

### [How to use Pivot Tables?](#)

Explore the Dynamic Reports on the

**You can look for Dynamic Reports on the following main entry points:**

1. [Holdings by Biological Status](#)
2. [Holdings by Health Status](#)
3. [Holdings by Continent-Country](#)

1

	A	B	C
1	FAO	Quarter	Accs
2	Yes	Qtr3	1,500
3	Yes	Qtr4	2,000
4	No	Qtr3	600
5	No	Qtr4	1,500
6	No	Qtr3	4,070
7	No	Qtr4	5,000
8	Yes	Qtr3	6,430

2

3

	E	F	G
	Sum of Dist	Quarter	
	FAO	Qtr3	Qtr4

Address <http://venus/appdb/research/bioinfo/cipinvi/templates/InvitroPropagations3.xls>

A1 = Health Status

	A	B	C	D	E	F	G
1	Health Status	All Health Status					
2	Camara	All Camara					
3	Time	All Time					
4	Propagator	All Propagator					
5	Activity	All Activity					
6							
7							
8	Medium Use I	Medium Sname	Data	Potato	RTA	Sweetpotato	
9	Conservation	MCB	Vessels			2862	
10			Accs			878	
11		MMC22	Vessels		108		
12			Accs		29		
13		MOC22	Vessels		42		
14			Accs		14		
15		MOC32	Vessels		1521		
16			Accs		473		
17		MSM33	Vessels		33		
18			Accs		7		
19		MU22	Vessels		180		
20			Accs		60		
21		MU32	Vessels		33		
22			Accs		11		
23		S22	Vessels	4488			
24			Accs	1604			

Sheet1 / Sheet2 /

# Data warehouse – examples solutions

1. <a href="#">Holdings by Biological Status</a>	Totals of accesions by biological status.
2. <a href="#">Holdings by Health Status</a>	Totals of accesions by health status.
3. <a href="#">Holdings by Continent-Country</a>	Totals of accesions by administrations (continent and country).
4. <a href="#">Distribution by Accession</a>	Specific information about accesions by region and country of distributed materials from CIP-Lima.
5. <a href="#">Distribution by Institution Type</a>	Distribution materials by accesions, crop, type institution (CGIAR centers, NARS, NGO, etc).
6. <a href="#">Distribution by Country</a>	Number of consignments by country and crop distributed by CIP-Lima.
7. <a href="#">Distribution by Biological Status</a>	Distribution materials by accesions, crop, type form and biological status.
8. <a href="#">Distribution in Invitro Form</a>	<b>Invitro</b> distributions from internal (CIP), national and international distributed materials.
9. <a href="#">Distribution by Region</a>	Distribution materials by region. It includes the number of approved requests by region and crop.
10. <a href="#">Morphology</a>	Morphology of native potato with status active.

# Software development practices

- Use of RUP documents for planning and documentation
- Visually Model Software: EclipseUML - Omondo
- Languages: VB/.NET; Java, PHP, R, SAS, (JavaScript, Delphi, Perl, C, C++, Fortran, Python)
- Standardized on Eclipse editors for Java, PHP, C, R
- Reuse a UI Component Framework: Eclipse RCP
- Use of Tortoise for CVS access for VB, etc
- Use of CVS and Subversion for all software development
- JIRA (Tracking Issue management), Confluence (Wiki)
- Evaluating: Tools for automating builds and software development quality control



# Outlook

- Stronger collaboration with ICIS team for better harmonization of tools developed at CIP for potato and sweetpotato data management
- Planning using GMS, GMSSRCH & SETGEN to integrate to CIP Systems
- Statistical quality control tools for process management
- Integration of ontologies for data capture and quality control