CIP – RIU selected activities

Information systems for crop and genebank data management

Edwin Rojas (Software Developer Chief),

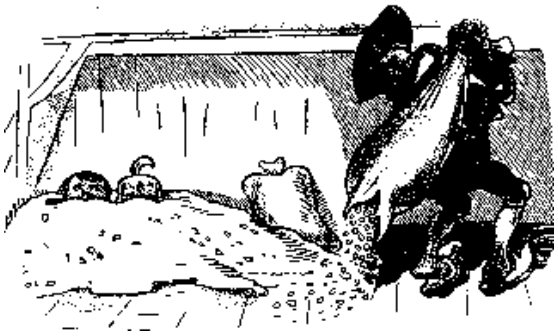
Reinhard Simon (RIU Head)

ICIS workshop 2006, CIMMYT

The Research Informatics story

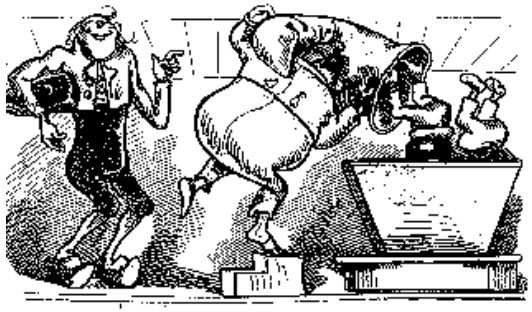
Research

Problematic situation

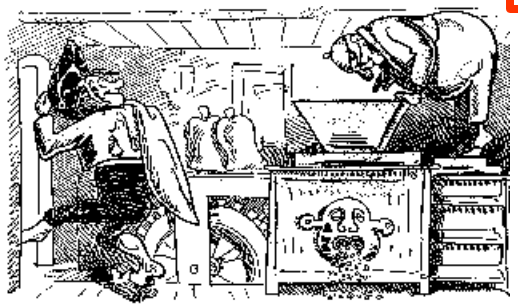


Hypothesis

Data capture



Data/information management



Analytical processing

Standardized reports

Research informatics

Publication for diverse uses/uptake

Presentation to CIP annual meeting 2004
Reinhard Simon

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

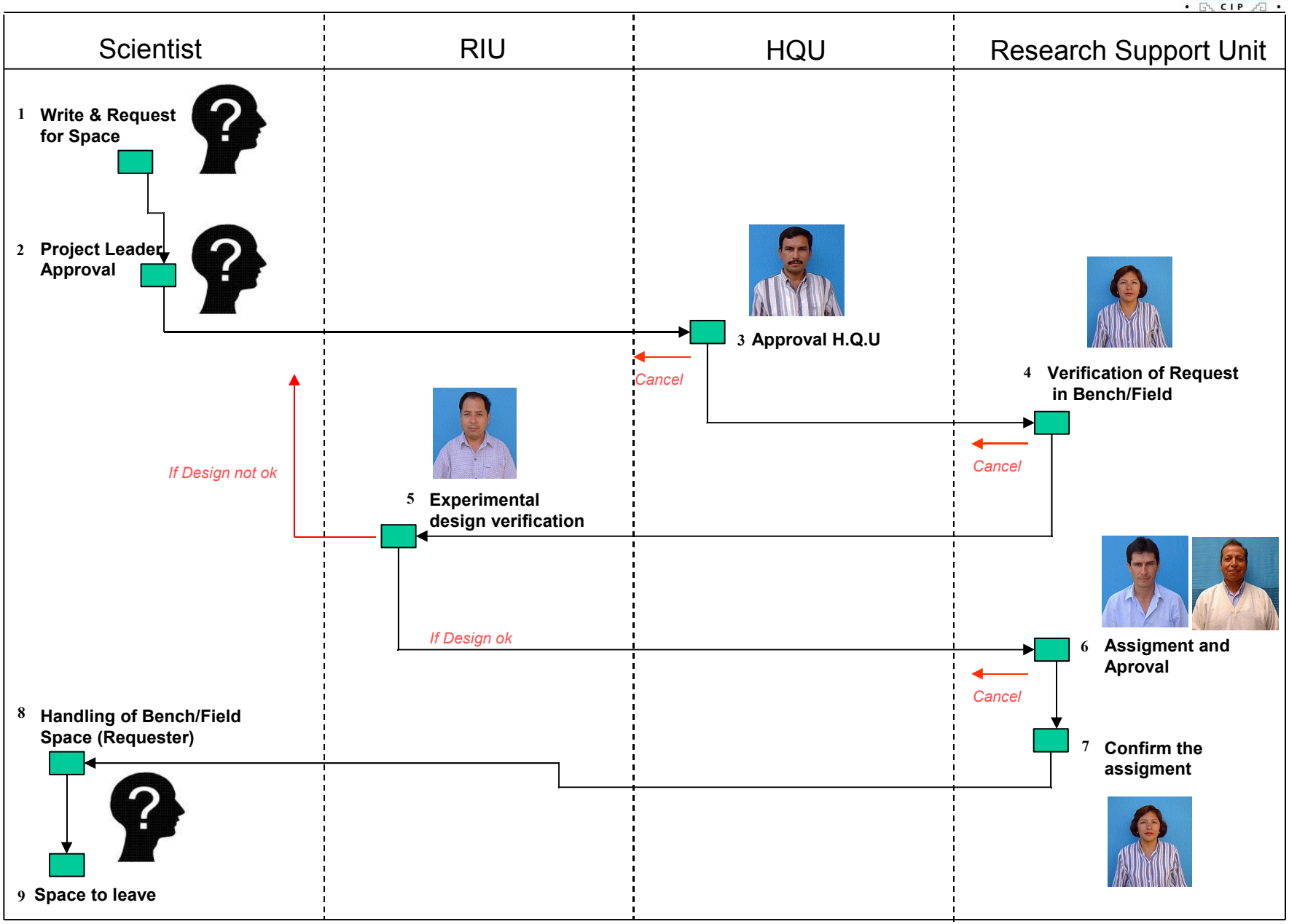
Selected RIU Activities

- 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
- CIPSTAT (Analysis Experiments) 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
- 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

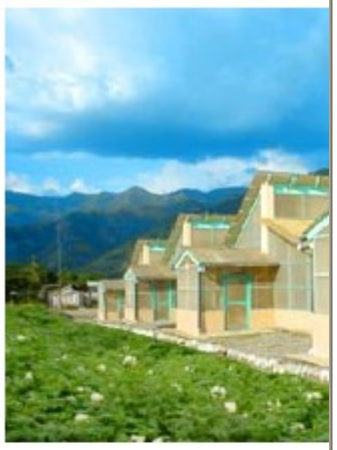


Date: 2004 / 03

Workflow Front-End: CIPSTATIONS

Experimental Station Management System (CIP-STATIONS)

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



Experimental Station Management S

Request for Field Space

Current Activity: Request Finalized

Have files: Yes

New | Save | Edit | Delete | Cancel | Files | Print | Follow up | Return

Year Number Next Activity

Search by number of Request Space

Year of request: 2004 Number of request: 51

Name Leader: LANDEO, JUAN Subleader: (empty)

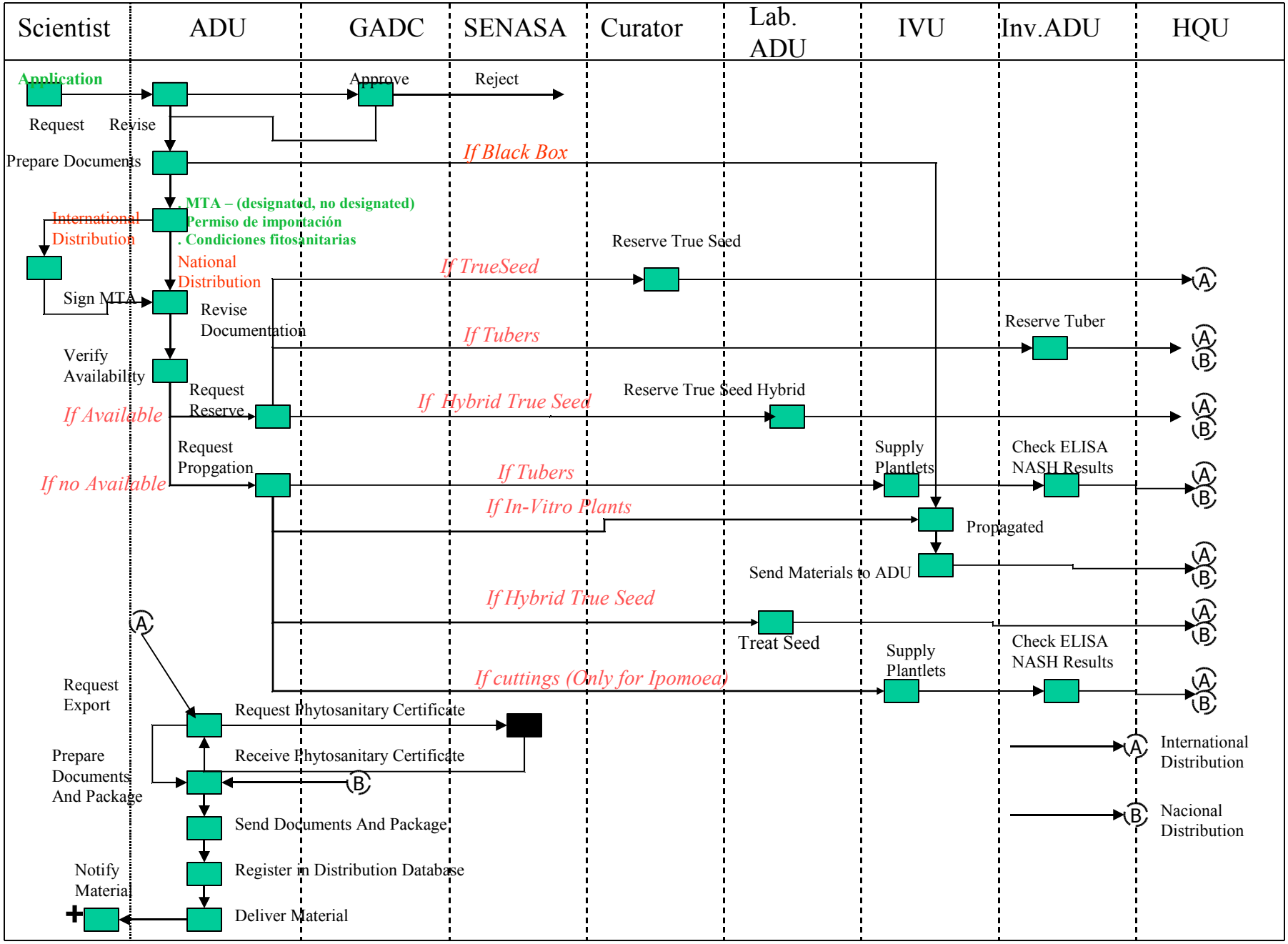
Name responsible approval request: BONIERBALE, MERIDETH Date request: 03/22/2004 mm/dd/yyyy

Name Assistant: GASTELO BENAVIDES, MANUEL ANT Project: 320203

Crop: Potato Other crop (empty) Type Use: ObservationPlot

Experimental Stations: San Ramon

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 PHQU) 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

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

We welcome any comments

Search a distribution of genetic material

2006 | Search | Export to Excel

All	Year of distribution :	All	
All	Destination:	All	
All	Status:	All	
All	Scientist :	All	
All	Institution:	All	

Search by number | Search old code

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

- 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

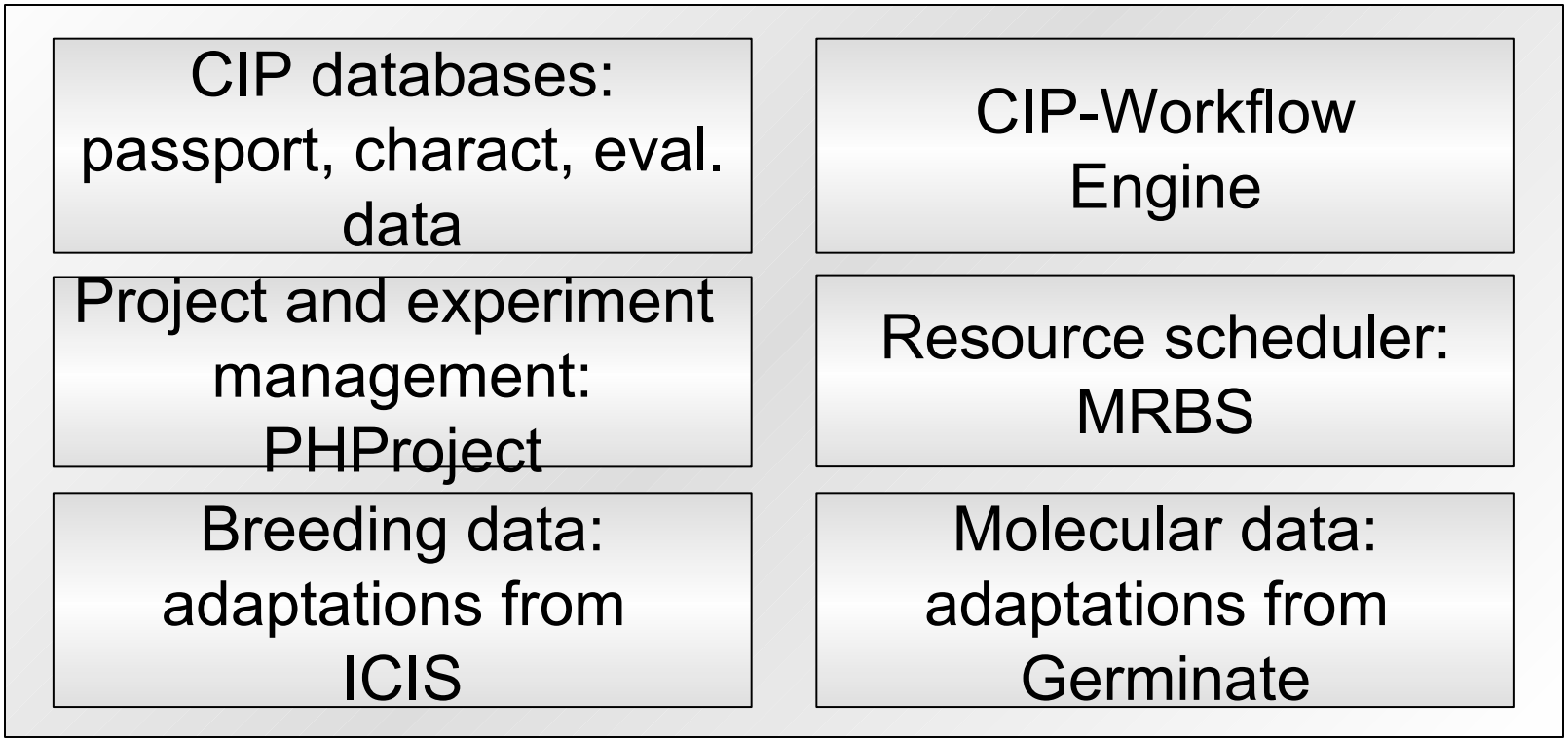


Diagram of the Experimental Design & Analysis Generator Module for the CIP-PEX System

PHPROJEKT

Register Experiment

Experimental Design

–RCBD
–Plot Design
–Lattice….

Input parameters

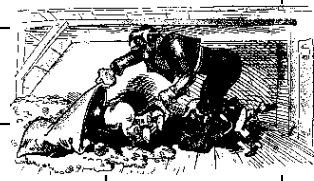
Add Factors

Add Variables

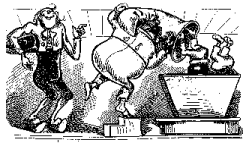
R

Experimental Map

Field Book

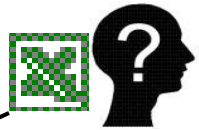


Data Capture

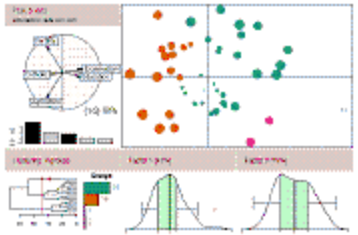


Central Database

MySQL



Analysis Modules



Statistical

GIS

Dataware House

DIVA-GIS

	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

	E	F	G
	Sum of Dist	Quarter	
	FAO	Qtr3	Qtr4
	Yes	7,930	2,000
	No	4,670	6,500
	Grand Total	12,600	8,500

Presentation to CIP annual meeting 2004
Reinhard Simon

CIP Mobile/Barcode Infrastructure Status 2006-01

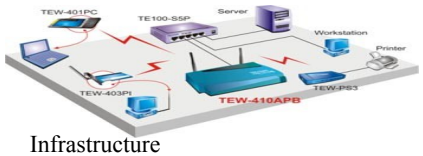
Barcode Pockets LAN Wireless

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

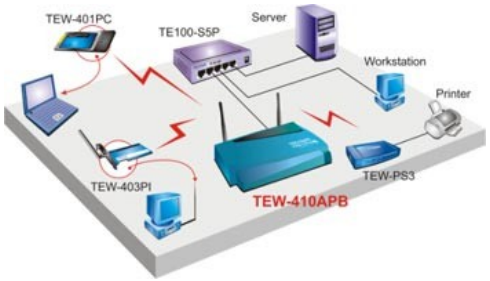
Applications

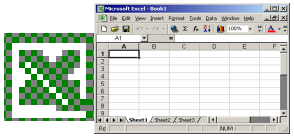
2004

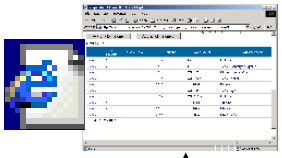
Improve Infrastructure and Wireless Covertures (from 11MB to 54 MB)

2005-2006

Environment Experiment Data Capture with Mobile Device







In-Vitro Genebank Performance Indicators

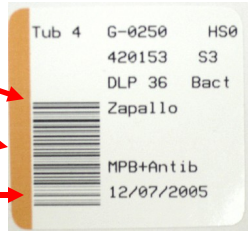
In-Vitro Genebank Activities	Barcode and Mobile Solution Technologies	Prior to Implementation Technologies
Locate a material by any staff	5 seconds	1800 seconds
Create a list for a “grid” of 15 entries	30 seconds	90 seconds
Print 90 labels for a “grid”	20 seconds	120 seconds
Clients request germplasm 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)

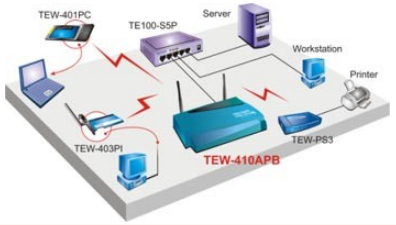
2. Barcode (1D)

Germplasm ID
+
Health Status
+
Culture Medium



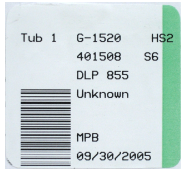
3. Thermal Printers

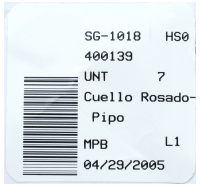
4. Wireless Networking

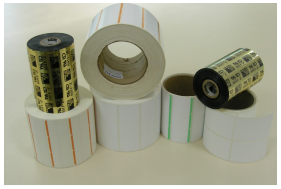


5. Hand Barcode Readers

6. Media: Labels and Ribbons







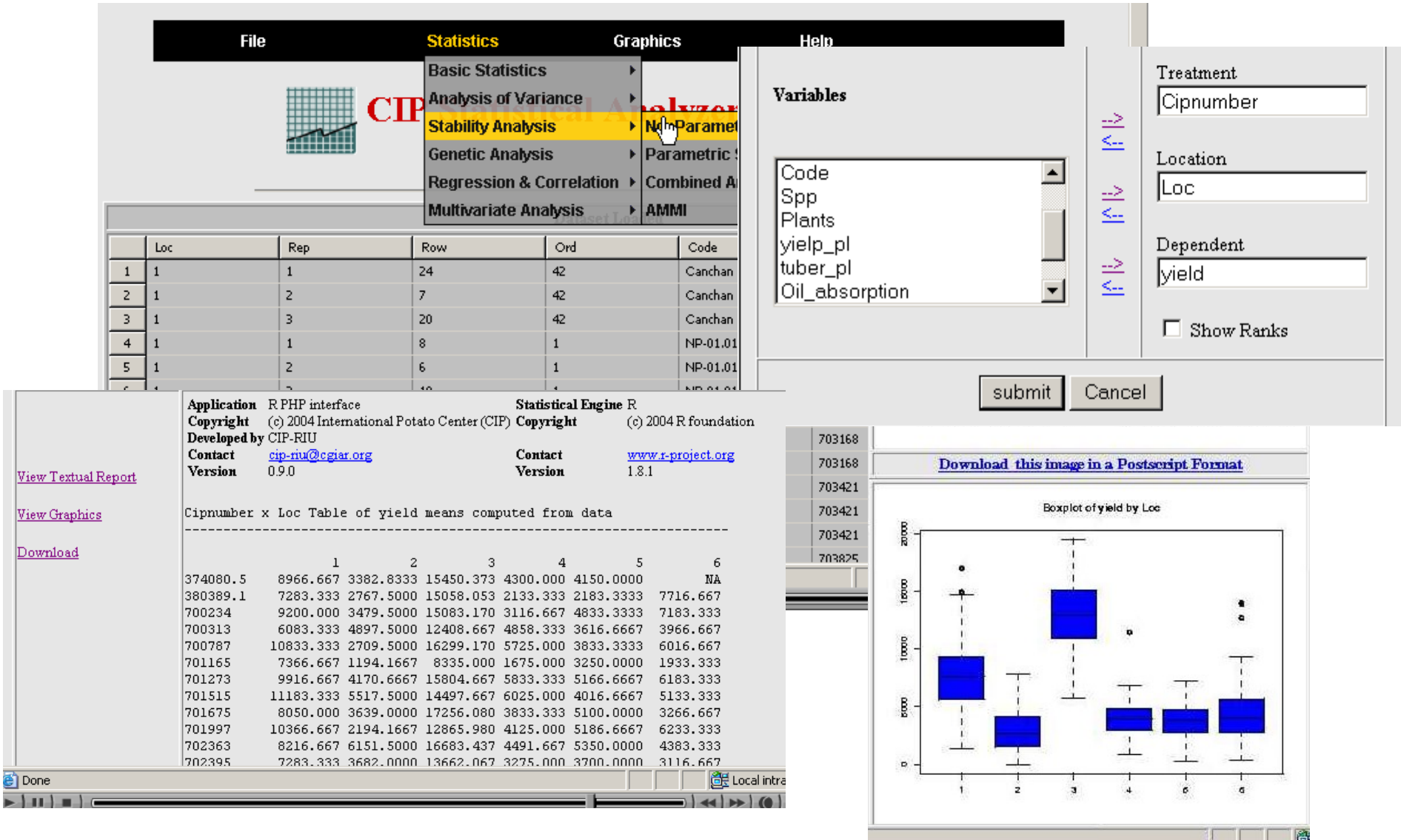
CIPSTAT

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

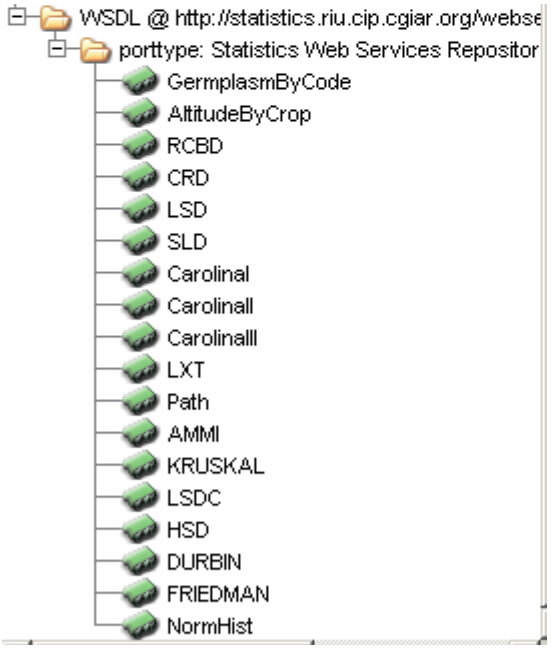
CIPSTAT – screenshot

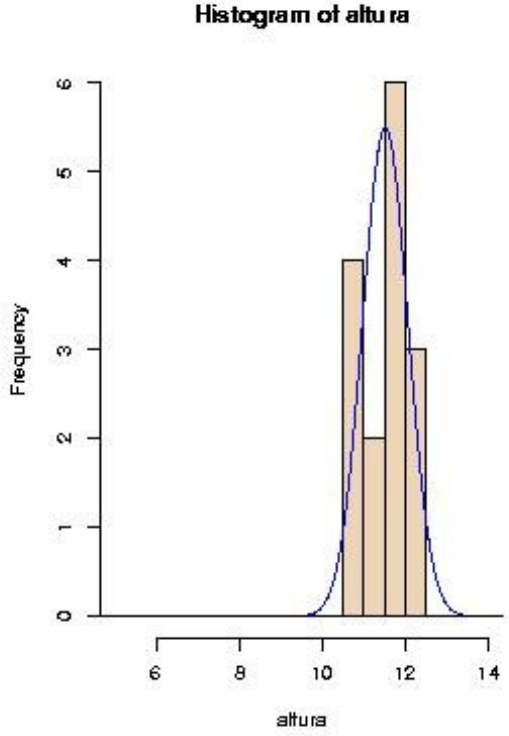
research.cip.cgiar.org/cipstat

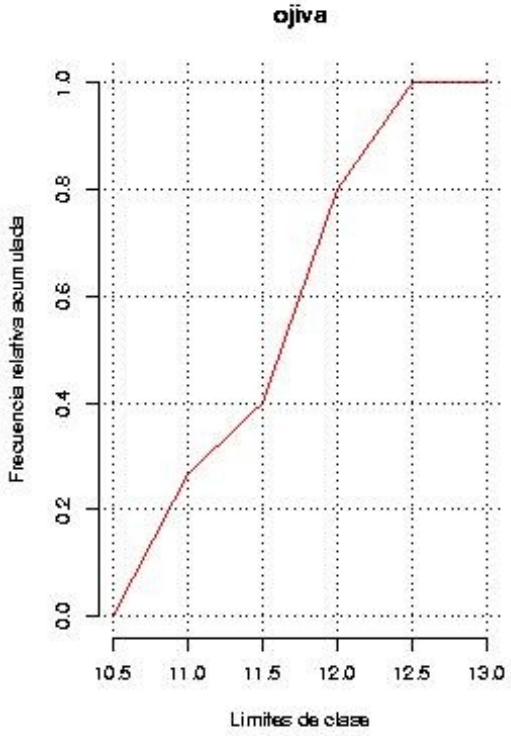


CIPSTAT – screenshot

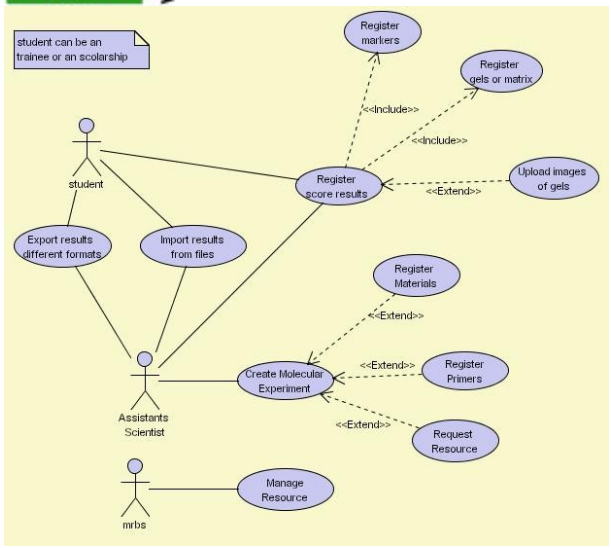
Interface for tools via WSDL – e.g. Taverna

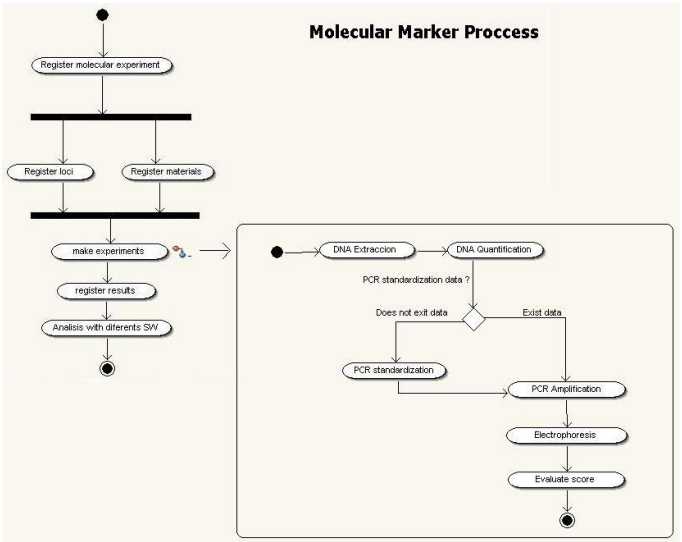


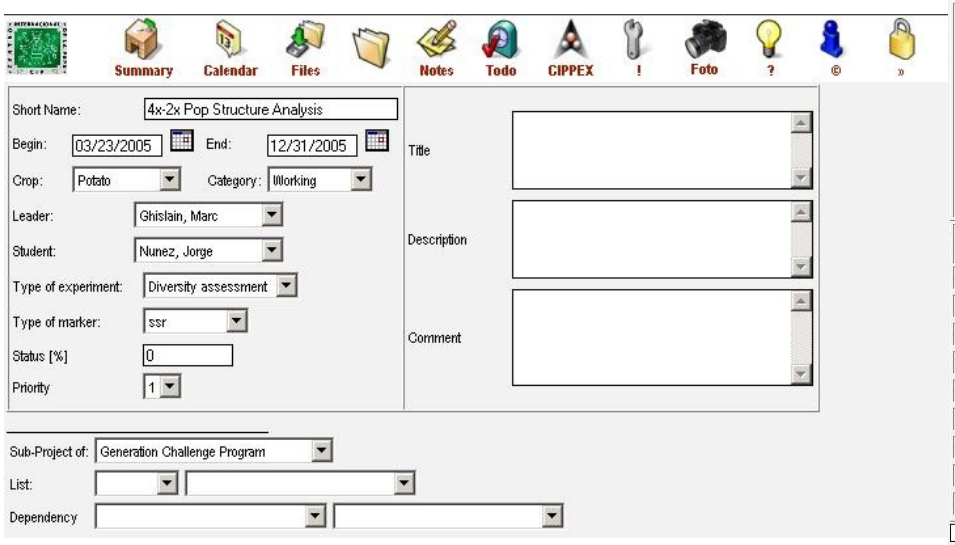


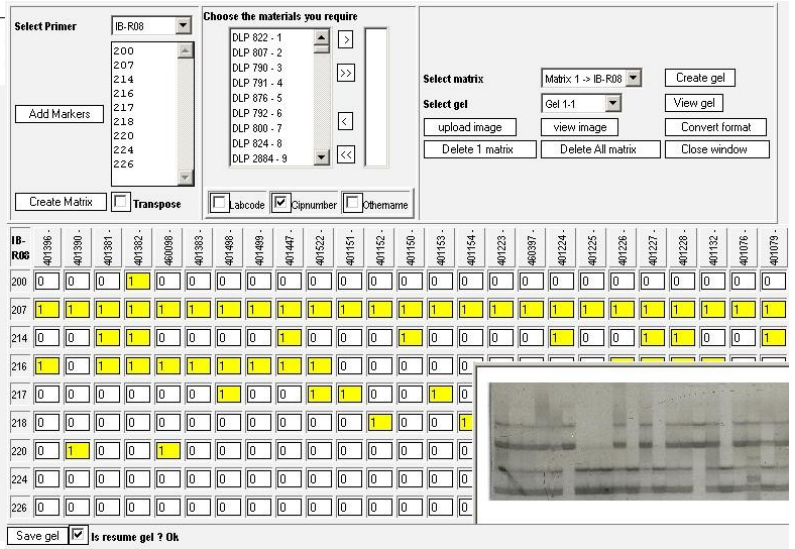


CIPPEX – LIMS molecular data

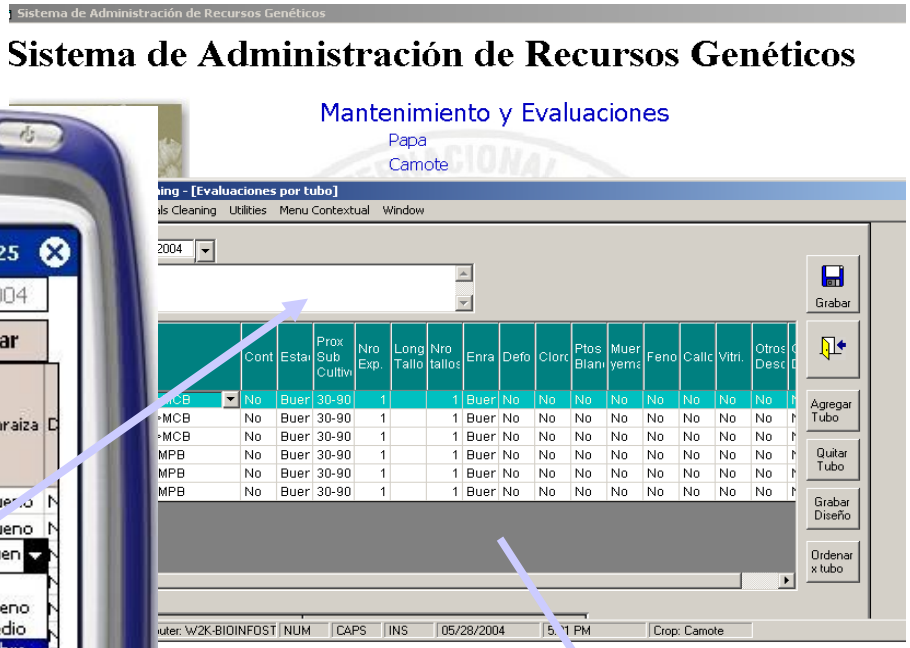


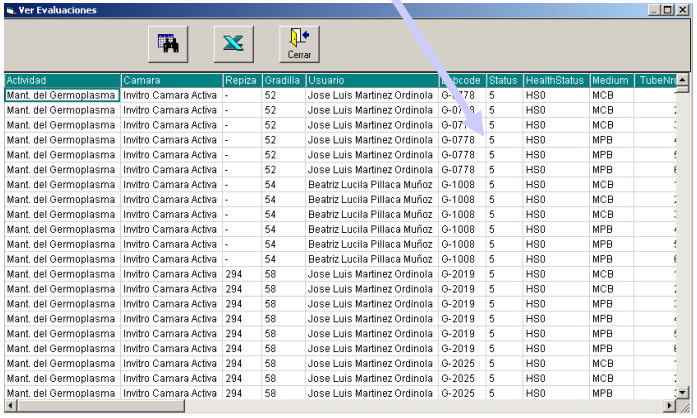






CIPTCL – LIMS for GR

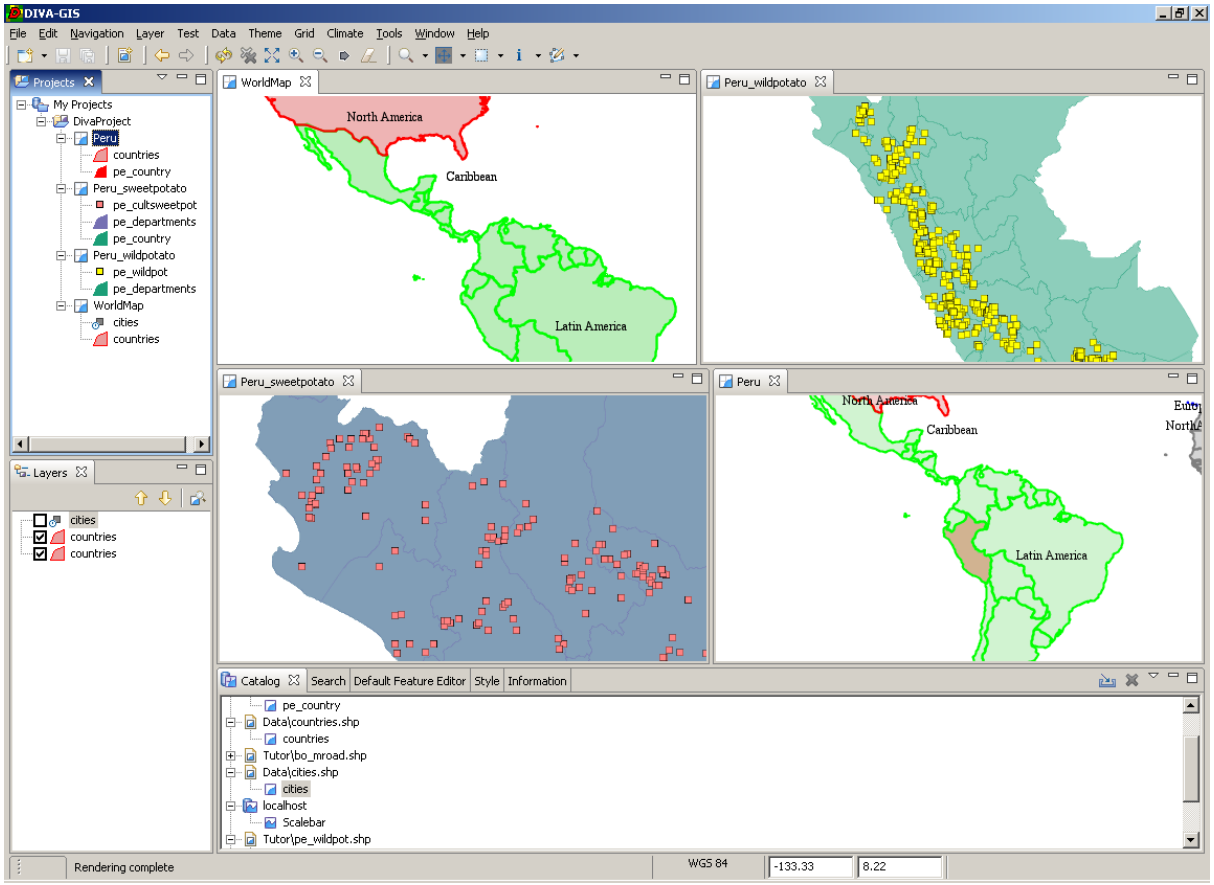


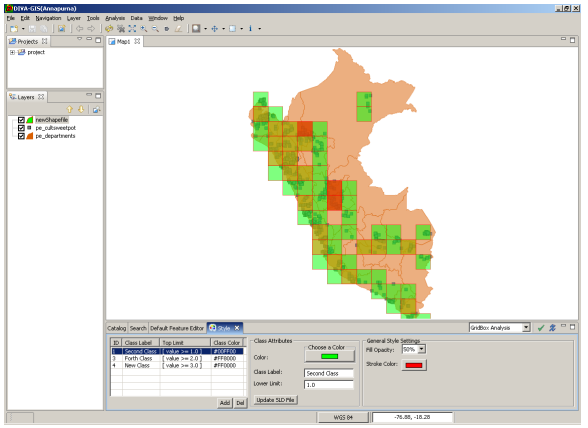


DIVA-GIS overview

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.

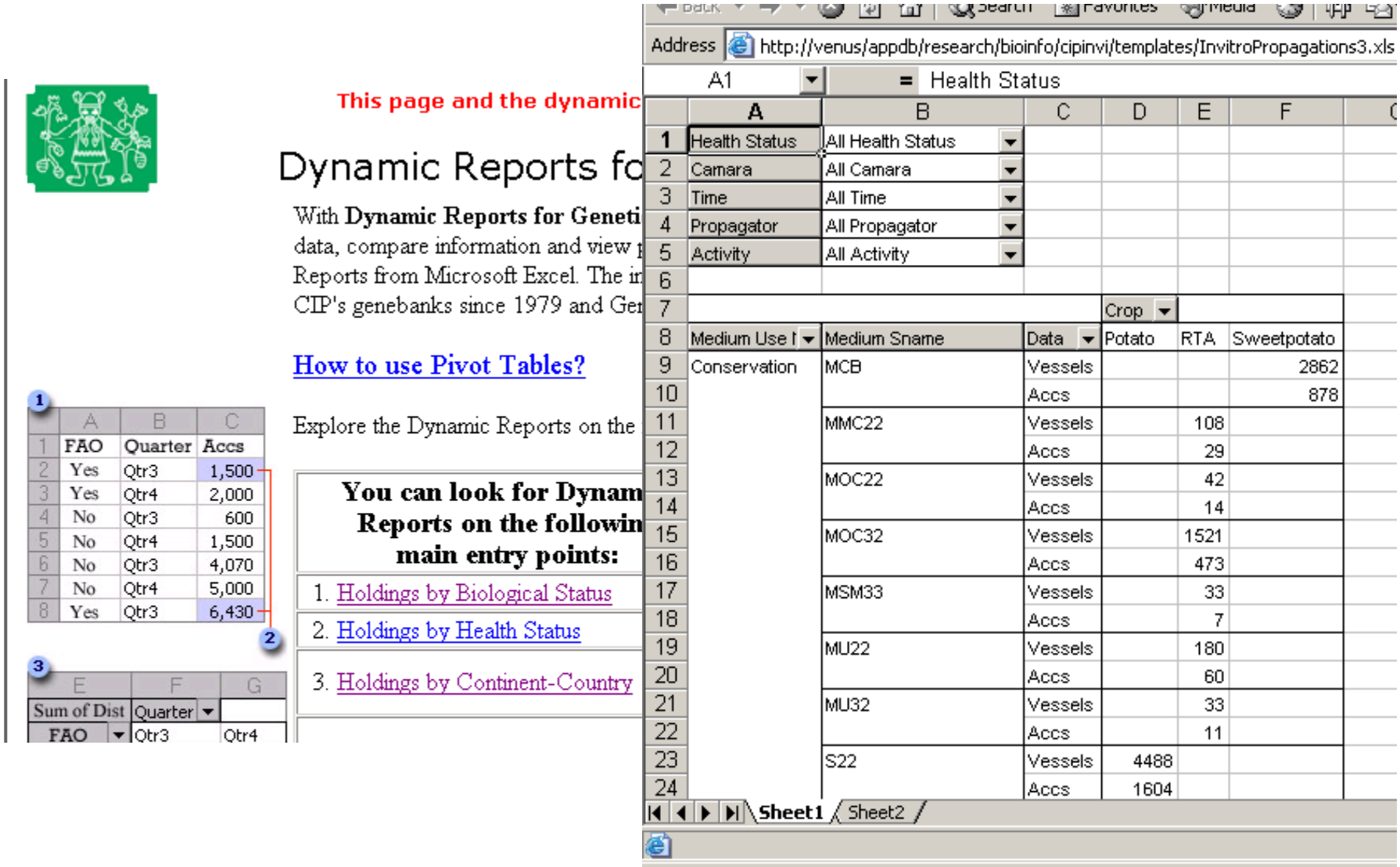




Data warehouse

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

Data warehouse – CIP Intranet



Data warehouse – examples solutions

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