

The Generation Challenge Programme Information Infrastructure

A quick overview of 2005 outputs

Tom Hazekamp (IPGRI)

**Compressed version of presentation to facilitate download*

Objectives of the Presentation

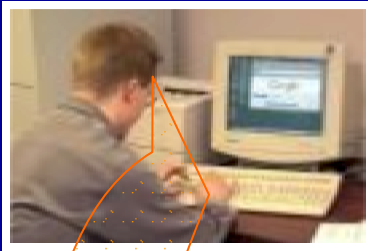
- ✓ **Provide an overview of the data flow strategy**
 - ✓ *Present a practical implementation at GCP partners location*
 - ✓ *Present the GCP Central Repository and its features to harvest GCP data sets*
- ✓ **Present the 2005 deliverables**
- ✓ **Highlight critical success factors**

Overview of the Challenges

The GCP Data infrastructure faces a diversity of complicating factors:

- ✓ **Standards/Models** (e.g. passport, genotype...)
- ✓ **Formats** (e.g. files, templates, databases...)
- ✓ **Availability** (e.g. off-line, on-line=distributed...)
- ✓ **Tracking** (e.g. different versions...)
- ✓ **Querying** (e.g. data compilation from multiple sources)
- ✓ **Visualizing** (e.g. aggregation, outputs for specific analytical tools etc...)
- ✓ **Access** (e.g. download options...)
- ✓ more...

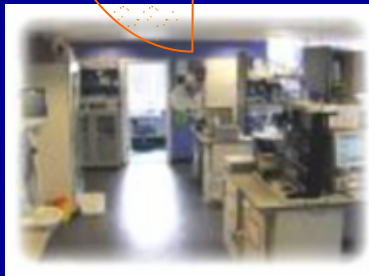
Implementation Strategy



A Central Repository that provides a “one stop shop” window to all GCP data sets



A robust GCP data infrastructure that harvests distributed data sources



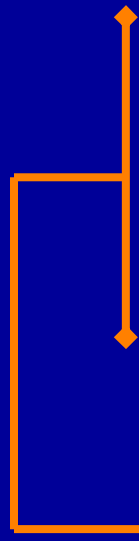
A distributed community of data providers located at different GCP sites

Harvesting should be easy...



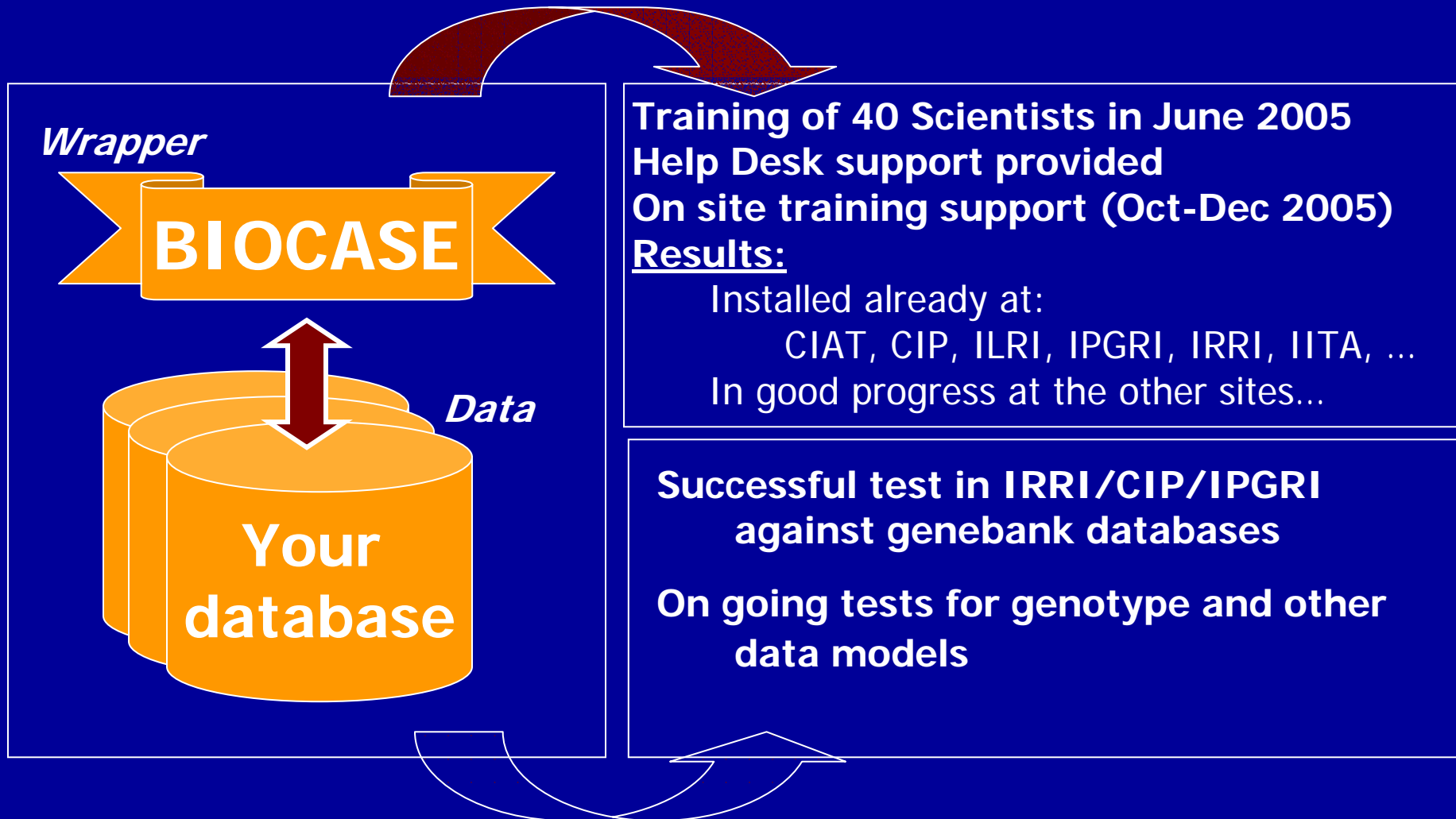
Data submission:

- ✓ **A simple mechanism for off-line data sets**
 - Classic file upload of any data types
 - Easy desktop software to format data against the GCP data models = **Templates**
- ✓ **A GCP data Center at every GCP sites** providing on-line access through web-services.
 - Data Center = Collecting point at your GCP site of all GCP standardized data sets.
 - Updated data are immediately visible and shared with other GCP partners.

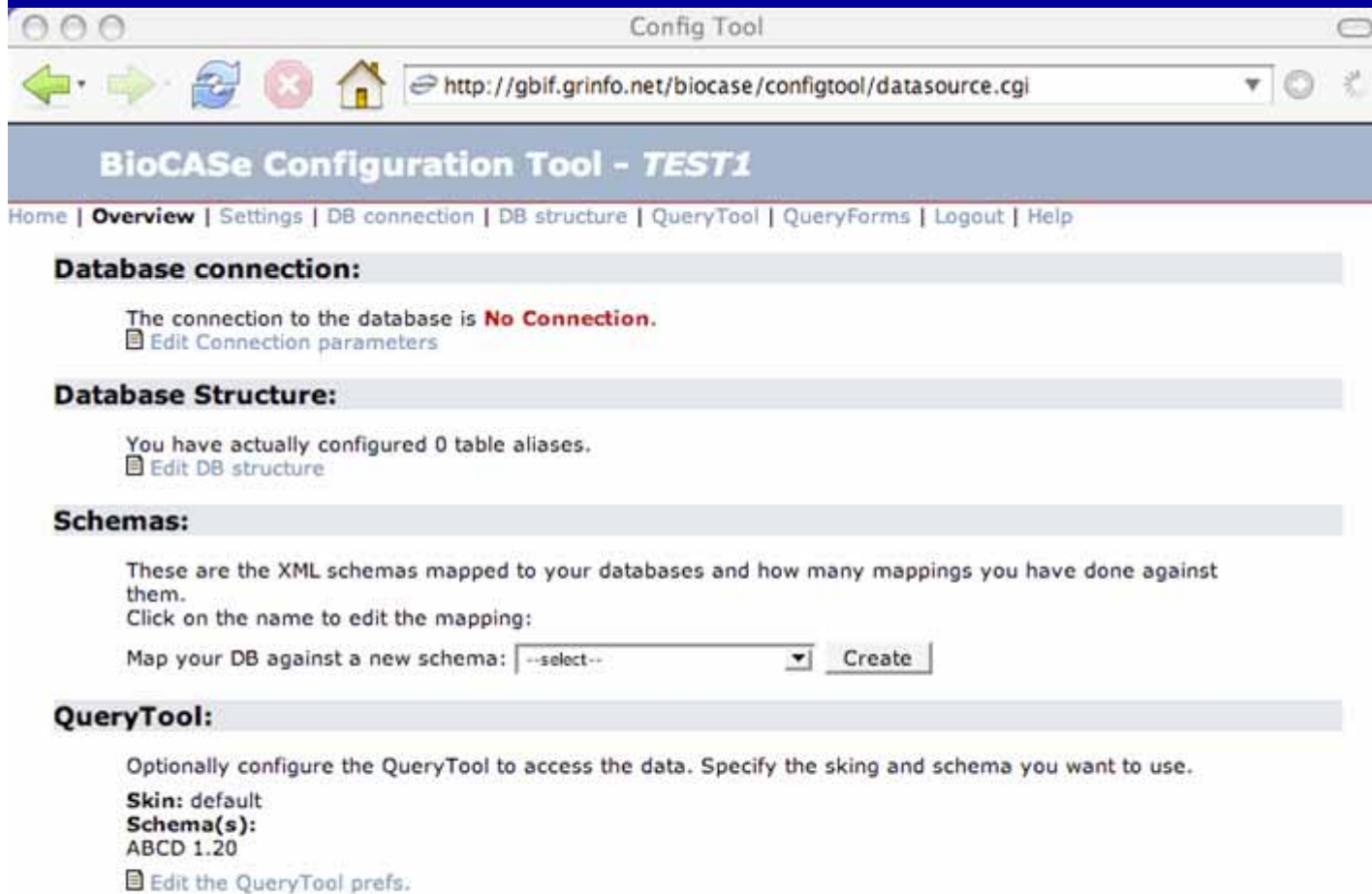


An optimum solution for the future

An optimum solution that works already !!!



Example of Deployment:



The screenshot shows a web browser window titled "Config Tool" with the address bar displaying "http://gbif.grinfo.net/biocase/configtool/datasource.cgi". The page header reads "BioCASE Configuration Tool - TEST1". A navigation menu includes "Home", "Overview", "Settings", "DB connection", "DB structure", "QueryTool", "QueryForms", "Logout", and "Help".

Database connection:

The connection to the database is **No Connection**.
[Edit Connection parameters](#)

Database Structure:

You have actually configured 0 table aliases.
[Edit DB structure](#)

Schemas:

These are the XML schemas mapped to your databases and how many mappings you have done against them.
Click on the name to edit the mapping:

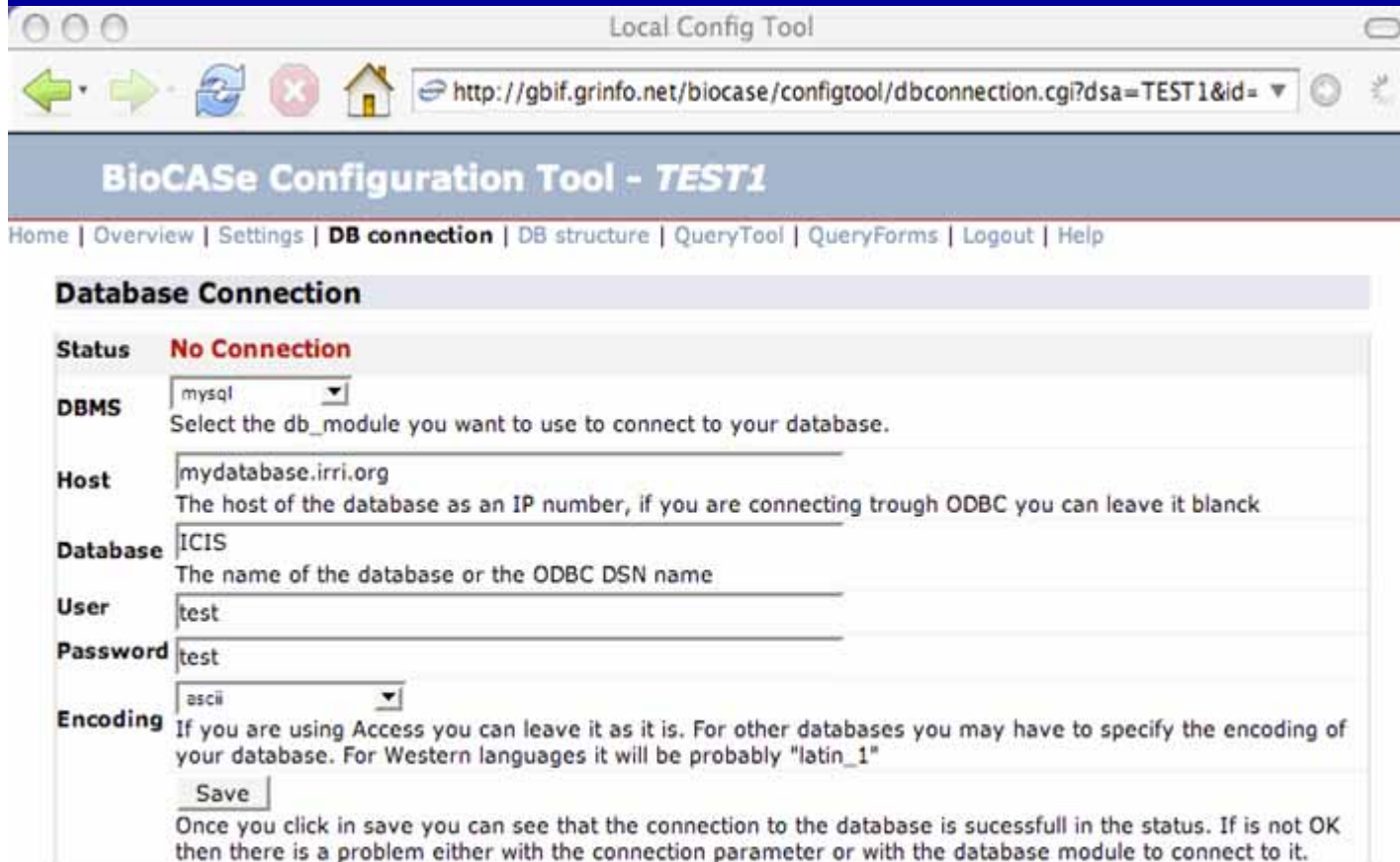
Map your DB against a new schema:

QueryTool:

Optionally configure the QueryTool to access the data. Specify the skin and schema you want to use.

Skin: default
Schema(s):
ABCD 1.20
[Edit the QueryTool prefs.](#)

Example of Deployment:



The screenshot shows a web browser window titled "Local Config Tool" with the URL <http://gbif.grinfo.net/biocase/configtool/dbconnection.cgi?dsa=TEST1&id=>. The page header reads "BioCASE Configuration Tool - TEST1" and includes a navigation menu: Home | Overview | Settings | **DB connection** | DB structure | QueryTool | QueryForms | Logout | Help.

Database Connection

Status **No Connection**

DBMS
Select the db_module you want to use to connect to your database.

Host
The host of the database as an IP number, if you are connecting trough ODBC you can leave it blanck

Database
The name of the database or the ODBC DSN name

User

Password

Encoding
If you are using Access you can leave it as it is. For other databases you may have to specify the encoding of your database. For Western languages it will be probably "latin_1"

Once you click in save you can see that the connection to the database is sucessfull in the status. If is not OK then there is a problem either with the connection parameter or with the database module to connect to it.

Example of Deployment:

The screenshot shows a web browser window titled "Local Config Tool" with the URL <http://gbif.grinfo.net/biocase/configtool/dbstructure.cgi?dsa=TEST&id=8078895669>. The page header is "BioCASE Configuration Tool - TEST" and includes a navigation menu: Home | Overview | Settings | DB connection | **DB structure** | QueryTool | QueryForms | Logout | Help.

Database Structure

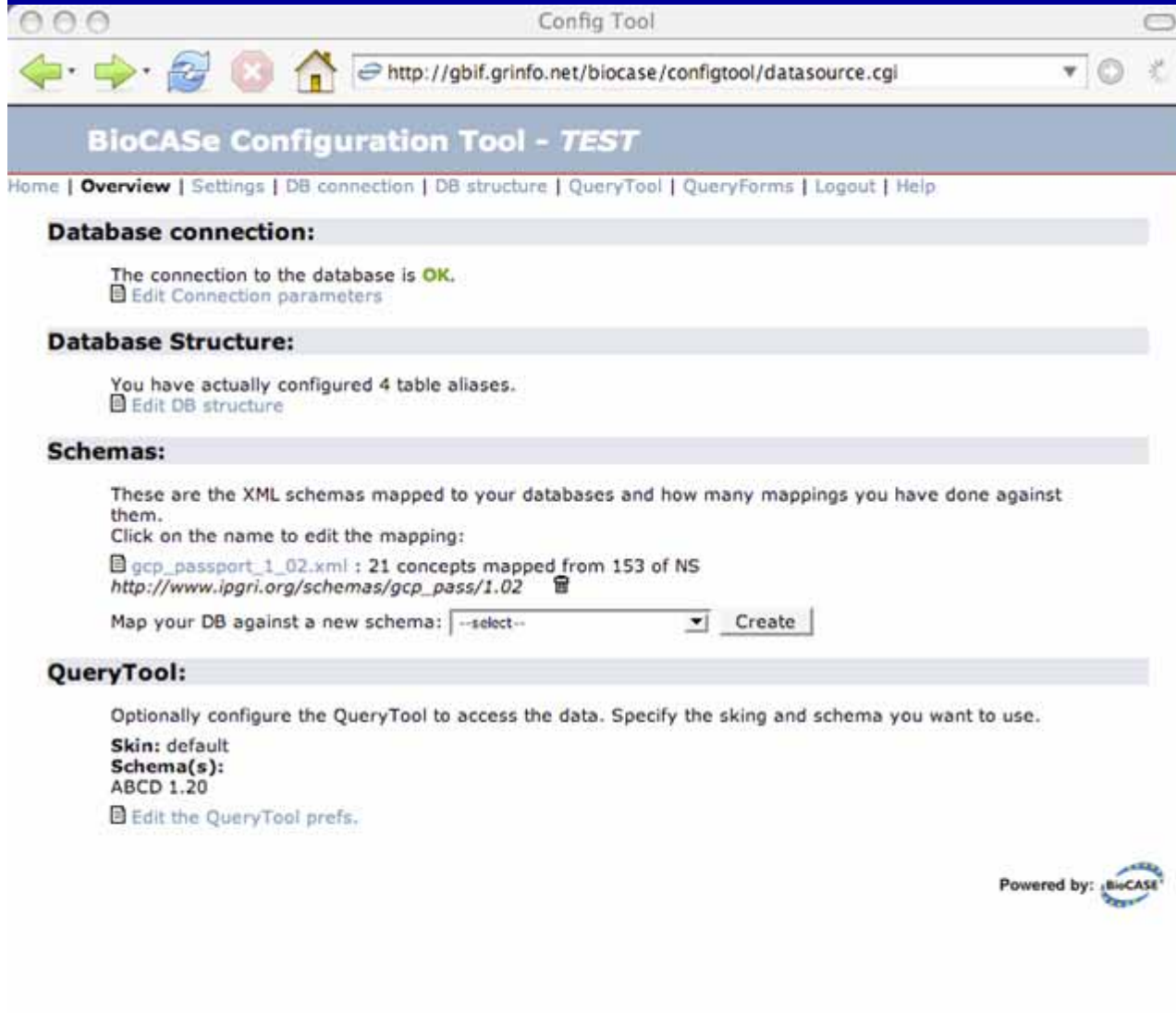
Use this tool to describe your database structure, so that the software knows how to query you database.

DB Connection status: OK

DB Structure status: Not implemented yet.

Alias	Table	Primary Key Attribute(s)	Foreign Key(s)
acc	acc	accnumb_ integer --new-- text	FK to alias col Attributes collcode_ integer --new-- text FK to alias inst Attributes instcode_ integer --new-- text FK to alias tax Attributes taxcode_ integer --new-- text new FK to alias --none--
col	col	code_ integer --new-- text	new FK to alias --none--
inst	inst	code_ integer --new-- text	new FK to alias --none--

Example of Deployment:



The screenshot shows a web browser window titled "Config Tool" with the address bar displaying "http://gbif.grinfo.net/biocase/configtool/datasource.cgi". The page content is as follows:

BioCASE Configuration Tool - TEST

Home | **Overview** | Settings | DB connection | DB structure | QueryTool | QueryForms | Logout | Help

Database connection:

The connection to the database is **OK**.
[Edit Connection parameters](#)

Database Structure:

You have actually configured 4 table aliases.
[Edit DB structure](#)

Schemas:


These are the XML schemas mapped to your databases and how many mappings you have done against them.
Click on the name to edit the mapping:
[gcp_passport_1_02.xml](#) : 21 concepts mapped from 153 of NS
http://www.ipgri.org/schemas/gcp_pass/1.02

Map your DB against a new schema:

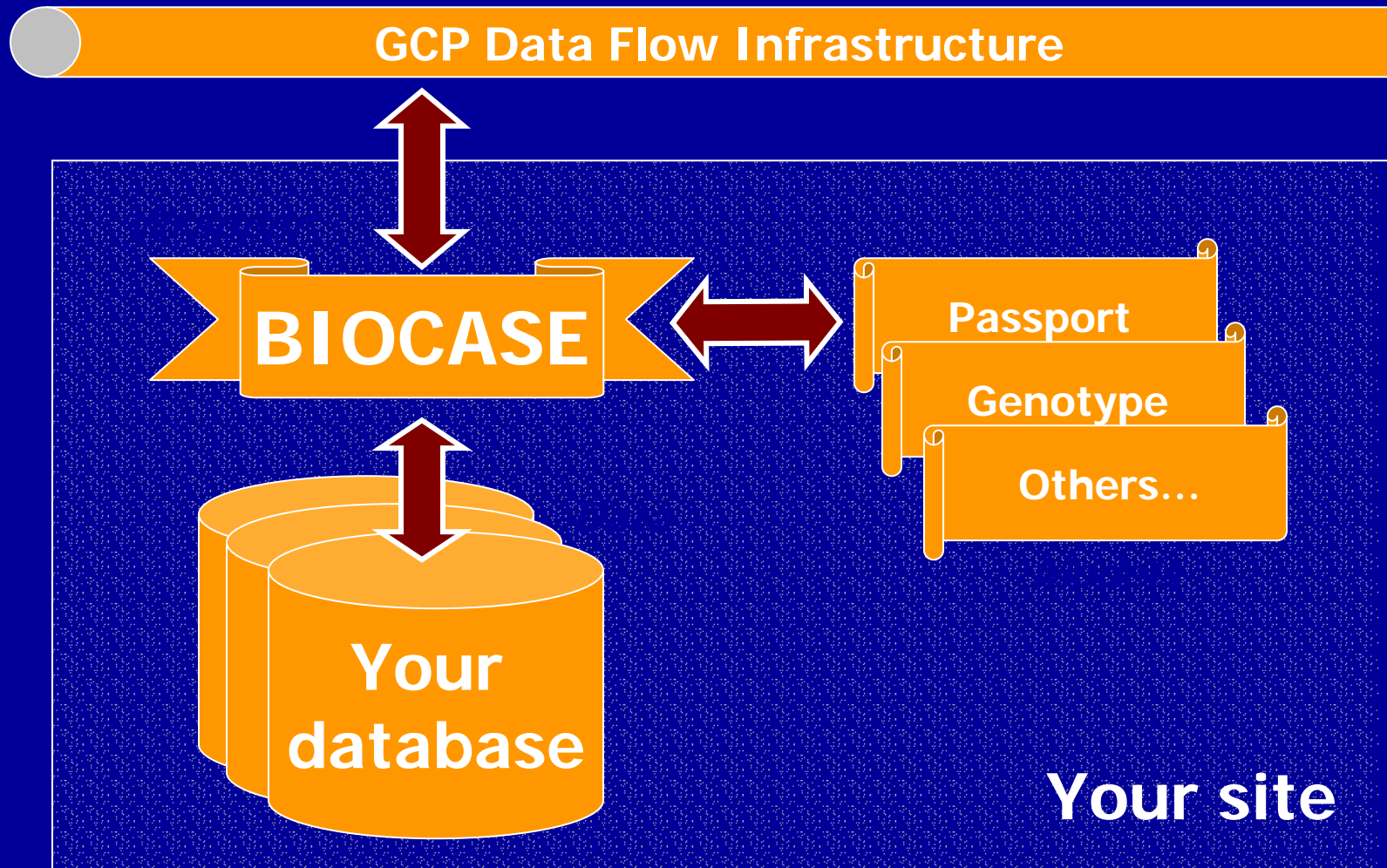
QueryTool:

Optionally configure the QueryTool to access the data. Specify the skin and schema you want to use.

Skin: default
Schema(s):
ABCD 1.20
[Edit the QueryTool prefs.](#)

Powered by: 

Example of Deployment:



PyWrapper Manual Query Form

Start » Utilities » QueryForms » Manual

Wrapper: <http://gbif.grinfo.net/biocase/pywrapper>

Debugging: off

Protocol:

```
<request>
<header><type>search</type>
</header>
<search>
<requestFormat>http://www.ipgri.org/schema
<responseFormat start='0' limit='10'>http:
<filter>
<like path='DataSets/DataSet/GermplasmSam
taxonomy/FullScientificName'>*SATIVA*</lik
</filter>
<count>>false</count>
</search>
</request>
```

Replace form with templates for a :
 ABCD scan, ABCD search, DWC Scan, DWC Search, Metaprofile Sc
 SPICE-1 Scan, SPICE-1 Search, SPICE-2 Scan, SPICE-2 Search

The BioCASE protocol filter operators

Comparison operators	unary	unbound	Logic
binary	isNull	in	binary
equals	isNotNull	and	and
notEquals		or	or
lessThan			
lessThanOrEquals			
greaterThan			
greaterThanOrEquals			
like			

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
<!--
XML generated by BioCASE PyWrapper software version 1.6.2 written by Markus Doering, Botanic Garden and Botani
-->
<response xsi:schemaLocation="http://www.biocase.org/schemas/protocol1.3 http://www.bgbm.org/biodivinf/schema/protocol1_3.xsd">
<header>
- <version software="Python Interpreter">
2.4.1 (#2, May 5 2005, 11:32:06) [GCC 3.3.5 (Debian 1:3.3.5-12)]
</version>
<version software="Wrapper">1.6.2</version>
<version software="DB module"/>
<version software="OS">posix</version>
<sendTime>2005-09-28T00:08:35+08:00</sendTime>
<source>66.102.11.99</source>
<destination>83.103.94.50</destination>
<type>search</type>
</header>
- <content recordDropped='10' recordCount='0' recordStart='0' totalSearchHits='11'>
- <DataSets>
- <DataSet>
- <GermplasmSamples>
- <GermplasmSample>
<GermplasmID>PHL001:Genebank:1</GermplasmID>
- <HoldingInstitute>
<FAOInstituteCode>PHL001</FAOInstituteCode>
- <NameOrganization>
International Rice Research Institute - Philippines
</NameOrganization>
<Street>Los Banos, Laguna</Street>
<CityAndState>Metro Manila</CityAndState>
<ZipCode>4030</ZipCode>
<Country>PHILIPPINES</Country>
<InstitutionalEmail>postmaster@irri.cgiar.org</InstitutionalEmail>
<Fax>+63 2-8450606</Fax>
<Ur>http://www.cgiar.org/irri</Ur>
<PrimaryContactName>Ruairadh Sackville Hamilton</PrimaryContactName>
</HoldingInstitute>
<CollectionName>IRRI-GRC</CollectionName>
<LocalUniqueID>1</LocalUniqueID>
<IsGenebankAccession>Y</IsGenebankAccession>
- <Classification>
- <CropNames>
<CropName>Rice</CropName>
</CropNames>
</CropNames>
- <Taxonomy>
<FullScientificName>O. SATIVA</FullScientificName>
<ScientificNameAuthor></ScientificNameAuthor>
<Genus></Genus>
<Species>O. SATIVA</Species>
<Cultivar>T 1242</Cultivar>
<CultivarGroup></CultivarGroup>
</Taxonomy>
</DataSets>
```


The Central Repository

GCP-CR Generation Challenge Programme Central Repository

http://gpcr.grinfo.net/index.php?app=file_upload

A CGIAR CHALLENGE PROGRAMME

Generation Challenge Programme

CULTIVATING PLANT DIVERSITY FOR THE RESOURCE POOR

home | about us | subprogramme 1 2 3 4 5 | resources | news | publications | virtual workspace | contact us | site map

GCP - Bioinformatics - Central Repository

[Start] [File Upload] [Search Files] [Import Data] [Search Data]

4 SUBPROGRAMME LEADER
Theo van Halbeek
Wageningen,
theo.vanhalle@wur.nl

Upload file to the Central Registry

* Select file to upload:

* Provider institute:

* Provider person:

* Provider email:

* Provided date: 2005-09-27 YYYY-MM-DD

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

* Source identifier:

* Source person:

Source date: YYYY-MM-DD

* Title:

Description:

* Type of resource: other:

Gcp subprogram:

Project activity:

Crop: other:

Keywords:

Status dataset: other:

Digital format: other:

CENTRAL REPOSITORY

Welcome to the Central Repository

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

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

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

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

[Central Repository Links](#)

[About the GCP-CR](#)

GENERAL LINKS

[Bioinformatics homepage](#)

[GCP Bioinformatics links](#)

Next steps in 2005

- ✓ *Finalize the deployment of all Data Centers*
- ✓ *Provide Help Desk support to GCP scientists*
- ✓ *Locate and Harvest all GCP data sources*
- ✓ *Initiate the indexing of all Data Sources*
- ✓ *Provide through the Central Repository a coherent view of all Data Sources*
- ✓ *Integrate the BioMOBY technology into a the Central Repository*

Critical success factors

- ✓ *Promote the GCP information sharing culture*
- ✓ *Ensure proper 'Data Transfer/Sharing Agreement'*
- ✓ *Ensure data flow and coordination from GCP Partner sites up to the central repository*
- ✓ *Promote linkages with non-GCP Partners*

A particular thanks to...

✓ *The BioMOBY Team:*

Mark Wilkinson, Ben Good...

✓ *The GBIF Team:*

Donald Hobern, Hannu Saarenmaa, Giorgos Ksouris

✓ *The BGBM Team:* *Berlin Botanical Garden*

Markus Doering, Javier De La Torre...

✓ *The Nordic Genebank (NGB):*

Bent Skovmand, Dag Terje Endresen

✓ *The CGIAR ICT-KM initiative:*

Enrica Porcari, David Balson...

✓ *The CGIAR SGRP:*

ICWG-GR, and SGRP Secretariat

✓ *as well as all the others....*

Thank you for your attention