

---

# Generation Challenge Programme Crop Models & Informatics Platform

---

**Richard Bruskwiech**

**Crop Research Informatics Laboratory  
International Rice Research Institute**





## Challenge Programme

“I **challenge** the next **generation** to use new scientific tools and techniques to address the problems that plague the world’s poor”

Dr. Norman Borlaug





## What is it?

- An international research programme established in 2003, projected to last 10 years, and hosted by the CGIAR with global partners from ARI and NARES
- Research Themes Directed to Crop Improvement:
  - Genomics and comparative biology across species
  - Characterization of genetic diversity for allele mining
  - Gene transfer technologies
- Five research subprogrammes, one of which is crop information systems development.



# Challenge Programme



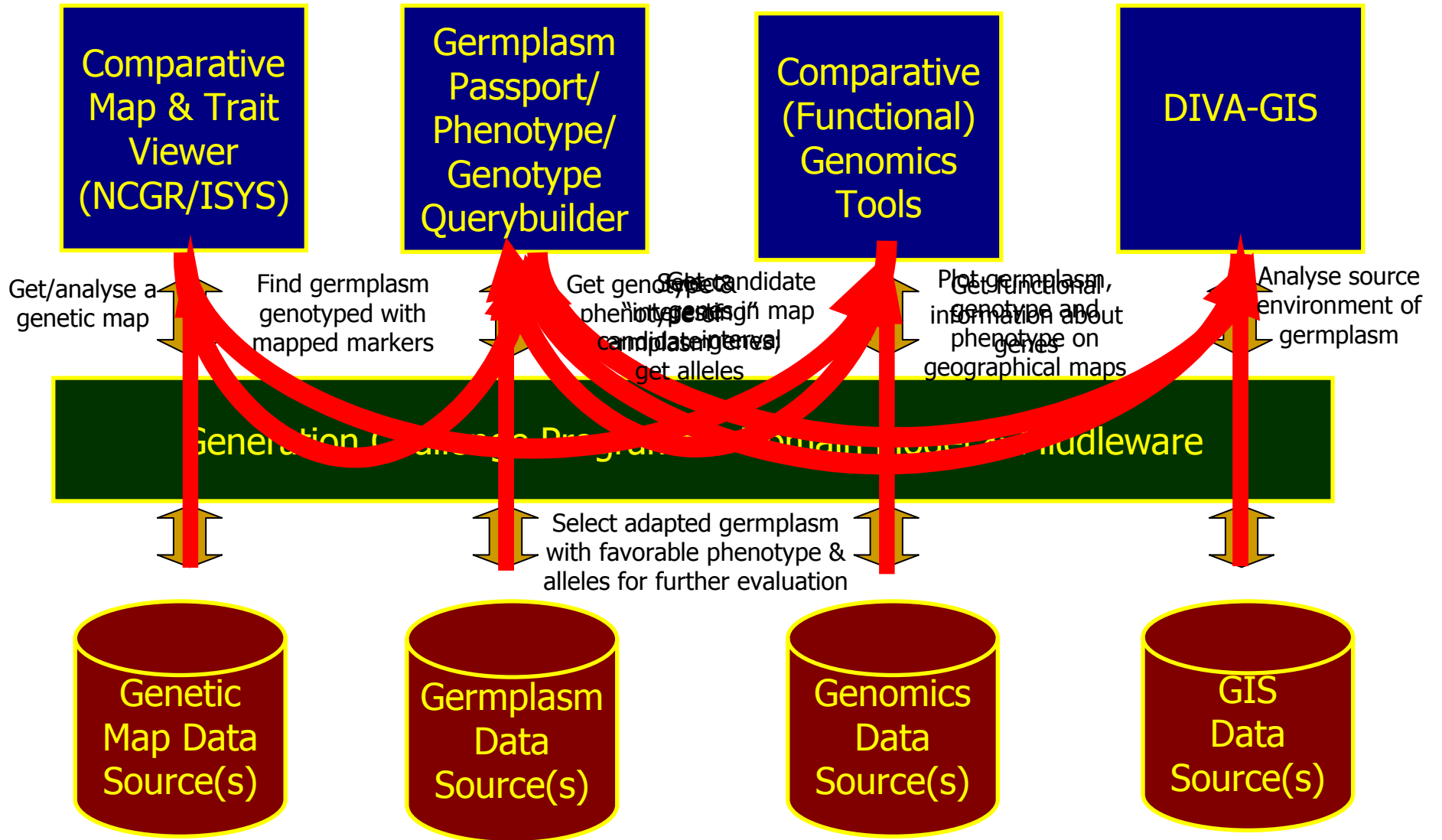
## Crop Information Systems: the Next



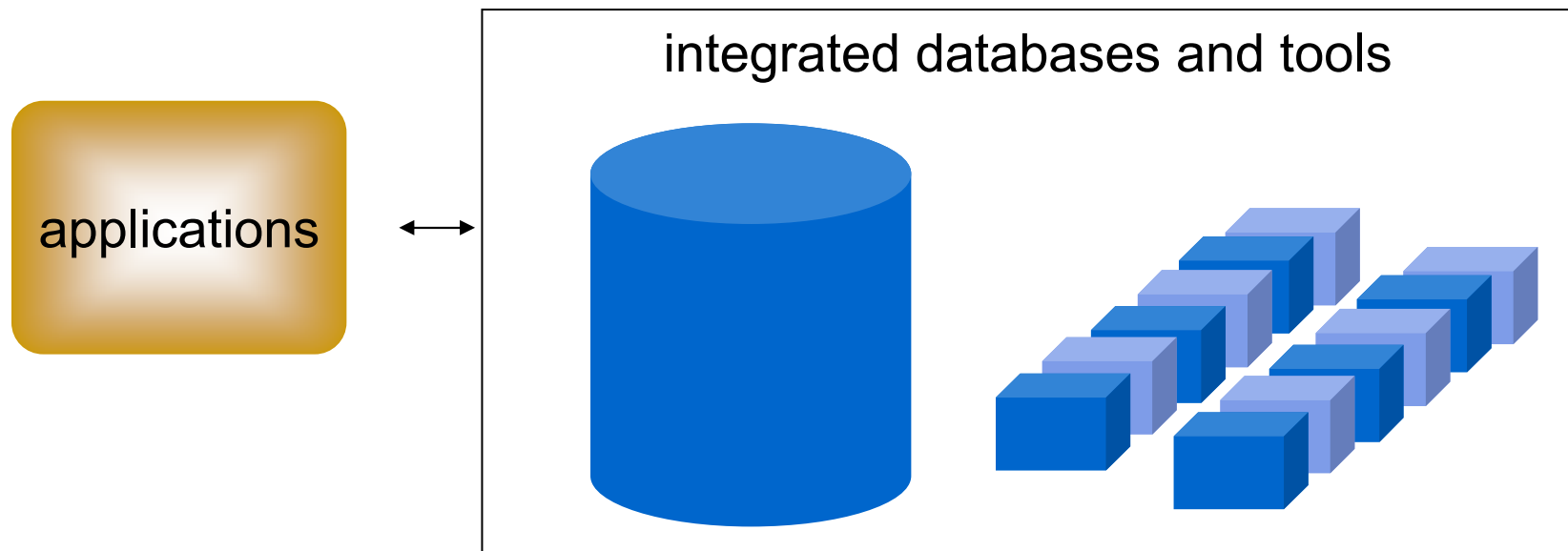
- Large, globally distributed consortium
- Diverse research requiring a diversity of tools
- Large data sets with diverse data types
- Many legacy informatics systems and tools
- Global data integration required...

**Key Issue: Interoperability**

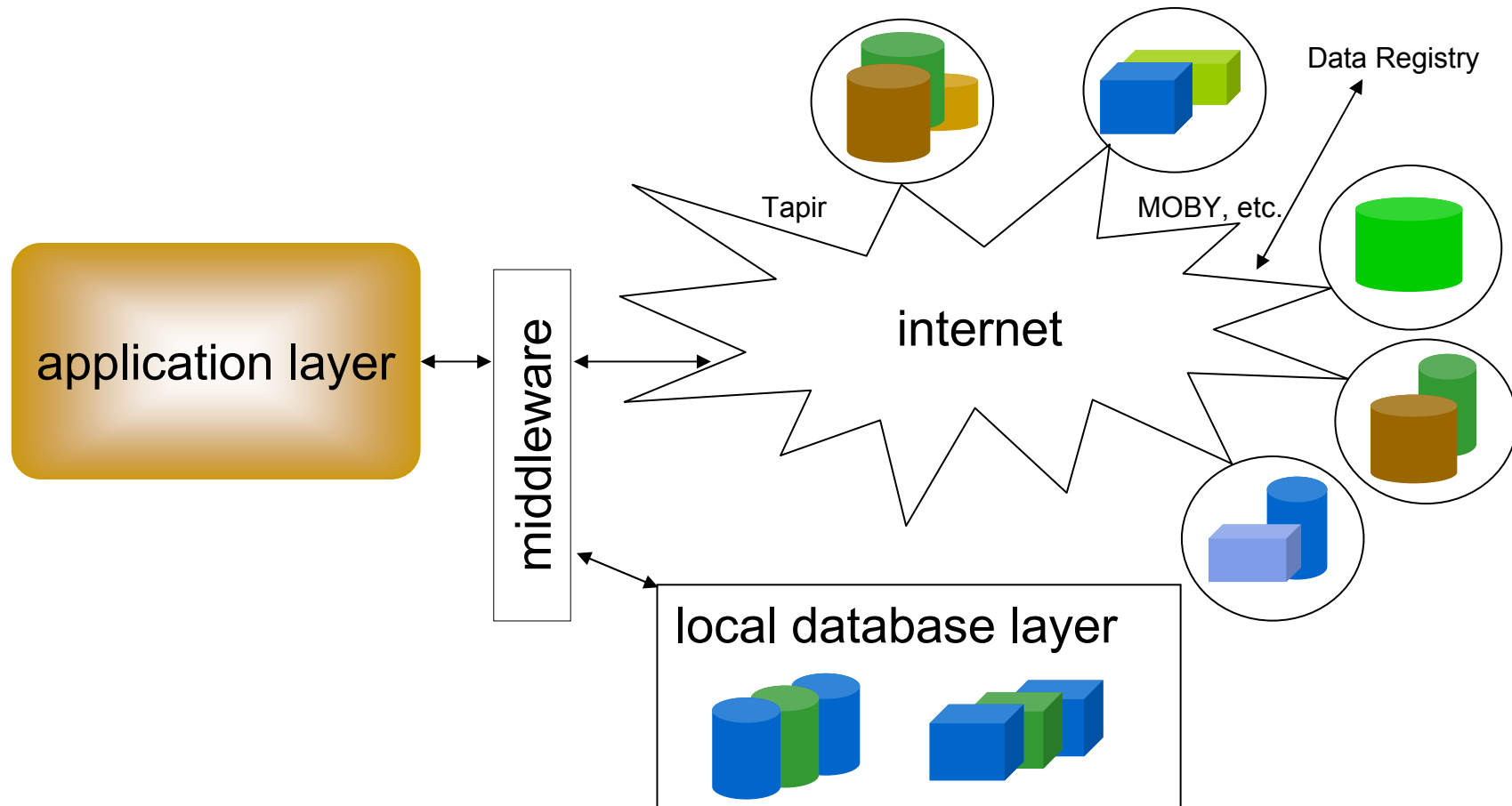
# A Generalized GCP Crop Research Integration Work Flow



An environment that provides improved access to data and analysis tools

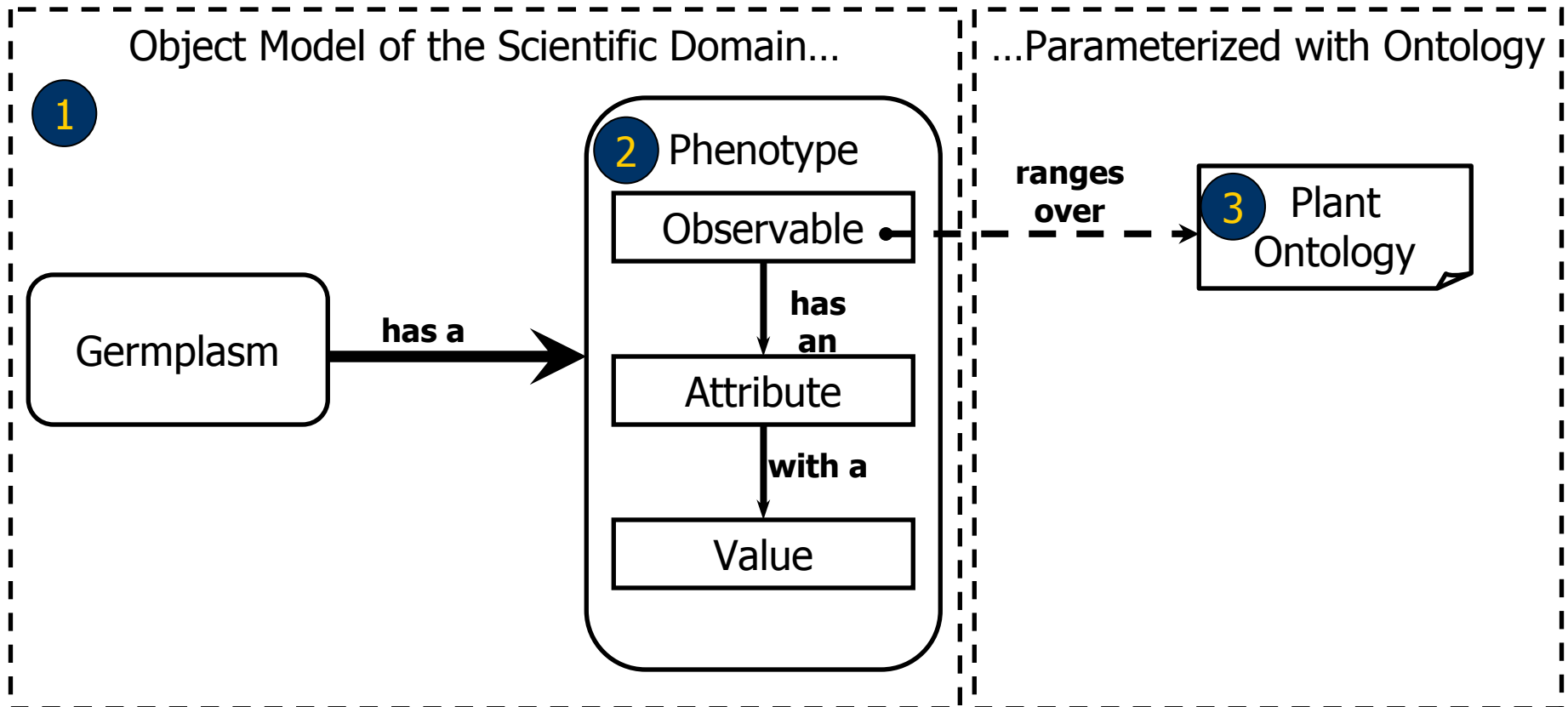


# GCP Information Platform – Developers’ Perspective





# Layers of Semantics



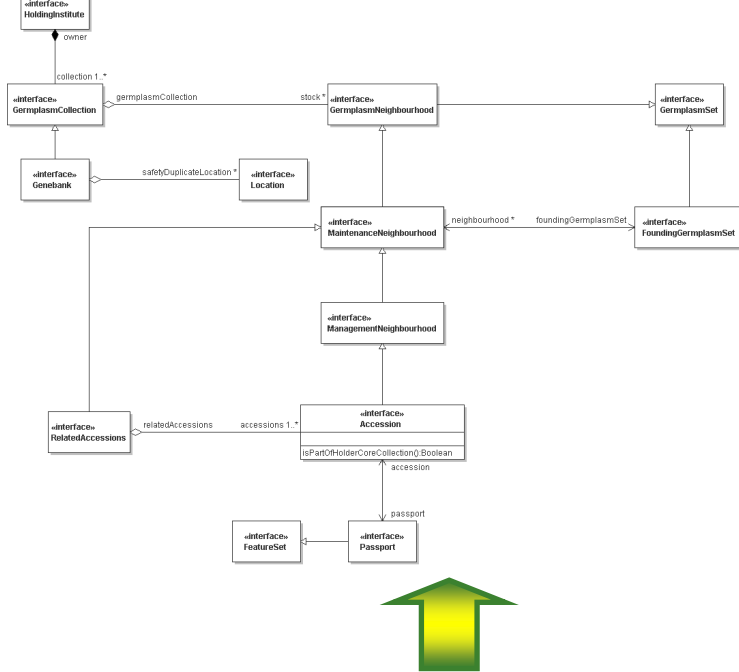


## Scope of GCP Domain Model & Ontology

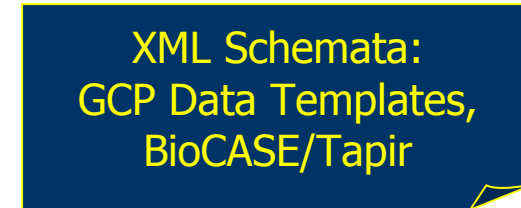
- **Core models:** generic concepts – identification, entities, features, organization, data management
  - Models heavily parameterized by ontology (e.g. entity and feature “type” attributes)
- **Scientific models:** extends core model into specific scientific scopes relevant to GCP:
  - Germplasm data (including genetic resources passport)
  - Genomics including genotypes, maps, sequences and functional annotation.
  - Phenotype data
  - Environmental data (including geographical location)

# GCP Domain Model Mappings onto Platform Specific Implementations

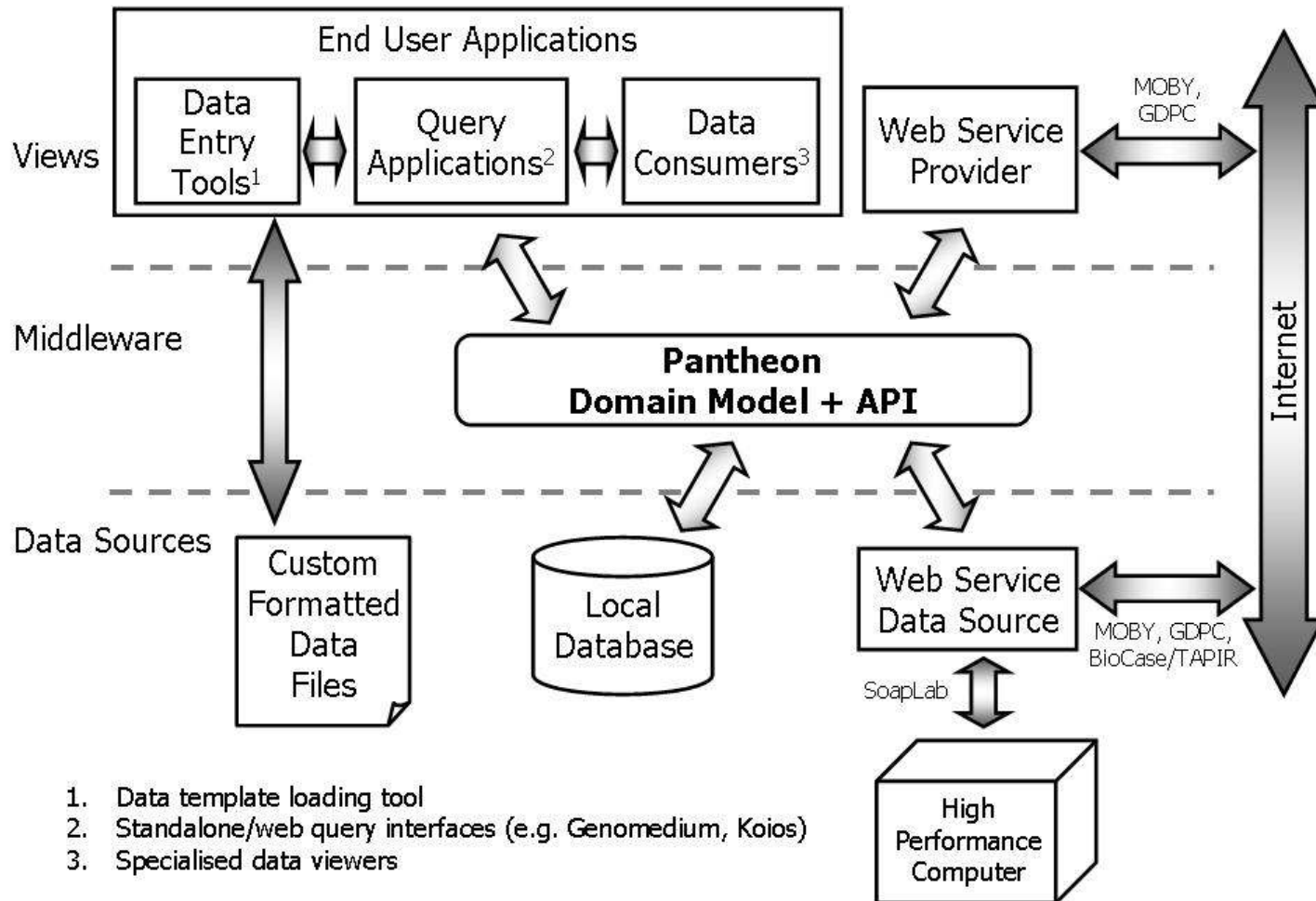
GCP Domain Model (UML/EMF)



SOAP Web Services  
(BioMOBY, SoapLab, GDPC)

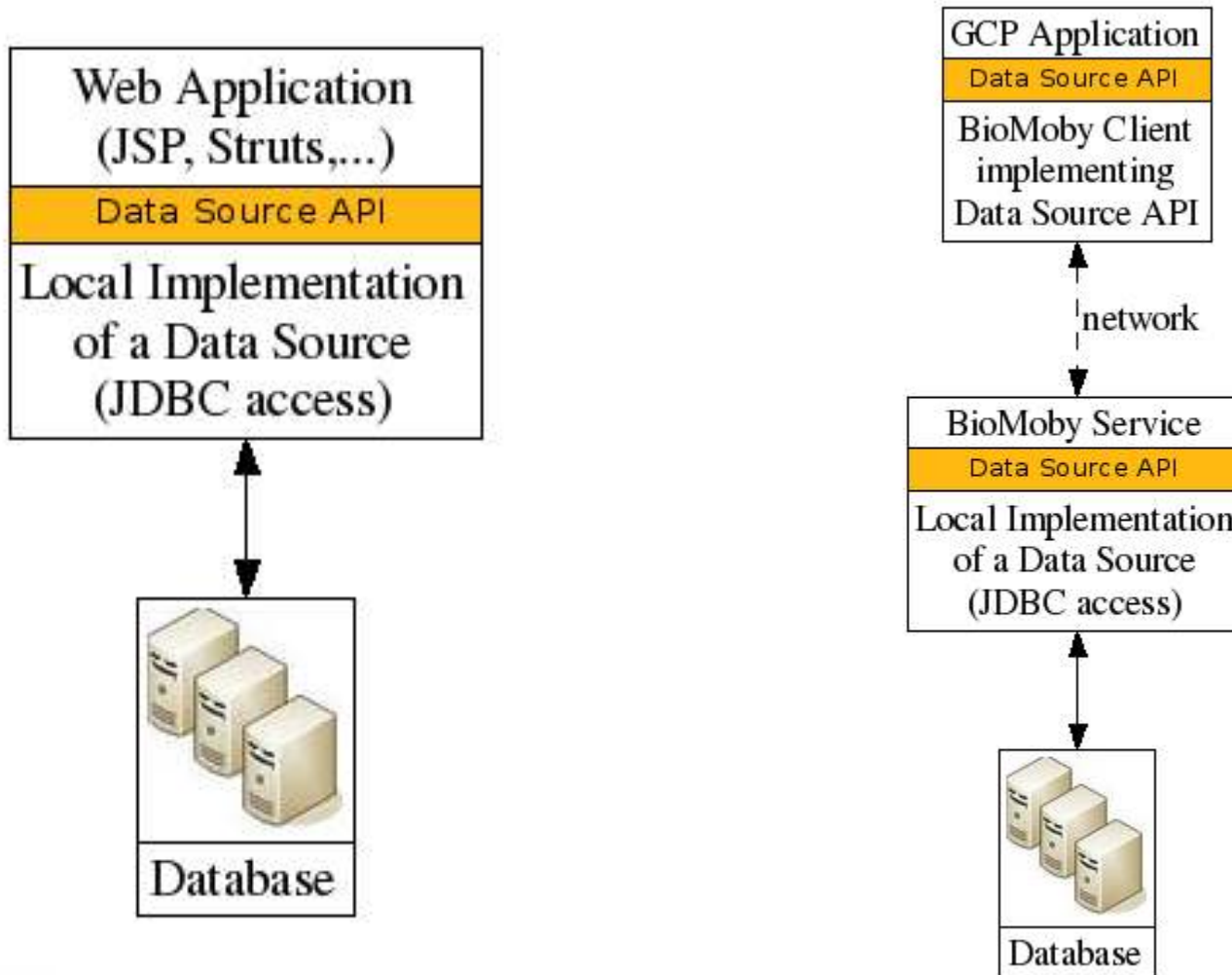


# Generation CP Platform



1. Data template loading tool
2. Standalone/web query interfaces (e.g. Genomedium, Koios)
3. Specialised data viewers

# GCP DataSources





## (Partial) Inventory of 3<sup>rd</sup> Party Data Resources targeted for wrapping as GCP Data Sources

Data Type	Description
<b>Microarray Data</b>	MAXD database with microarray datasets from diverse GCP commissioned or competitive projects.
<b>Genetic and QTL Mapping Data</b>	QTL data available in ICIS, TropGenes. Genomic Diversity and Phenotype Connector (GDPC) connecting to Gramene, Panzea, GrainGenes <i>et al.</i>
<b>Genomic Sequence Data and Annotation</b>	NIAS KOME full length cDNA and RAP genome databases (?), connected to GCP web services by NIAS. OryzaSNP and GCP comparative genomic databases. Public sequence databases (via BioJava?)
<b>Functional Genomics</b>	OryGenesDb mutant data (CIRAD); IR64 rice mutant database (IRRI); Tos17 database (NIAS).
<b>Germplasm Sample Characterization Data</b>	Germplasm, passport, genotype and associated field data available in ICIS databases; TropGenes, MGIS, ICRIS.



## GCP Platform Implementations

- Standalone workbench (“GenoMedium”)
  - Eclipse Rich Client Platform (RCP)
- Web-based workbench (“Koios”)
  - AJAX, PHP, Java (server side), Java Web Start
- NCGR Integrated SYStem (ISYS)
- Direct tool integration (e.g. GCP MaxdLoad)



# GCP Web-Based Search Engine

Summary of query hits

GCP semantics defined query

List of items matched

View details at 3<sup>rd</sup> party web site or in locally invoked 3<sup>rd</sup> party data viewer



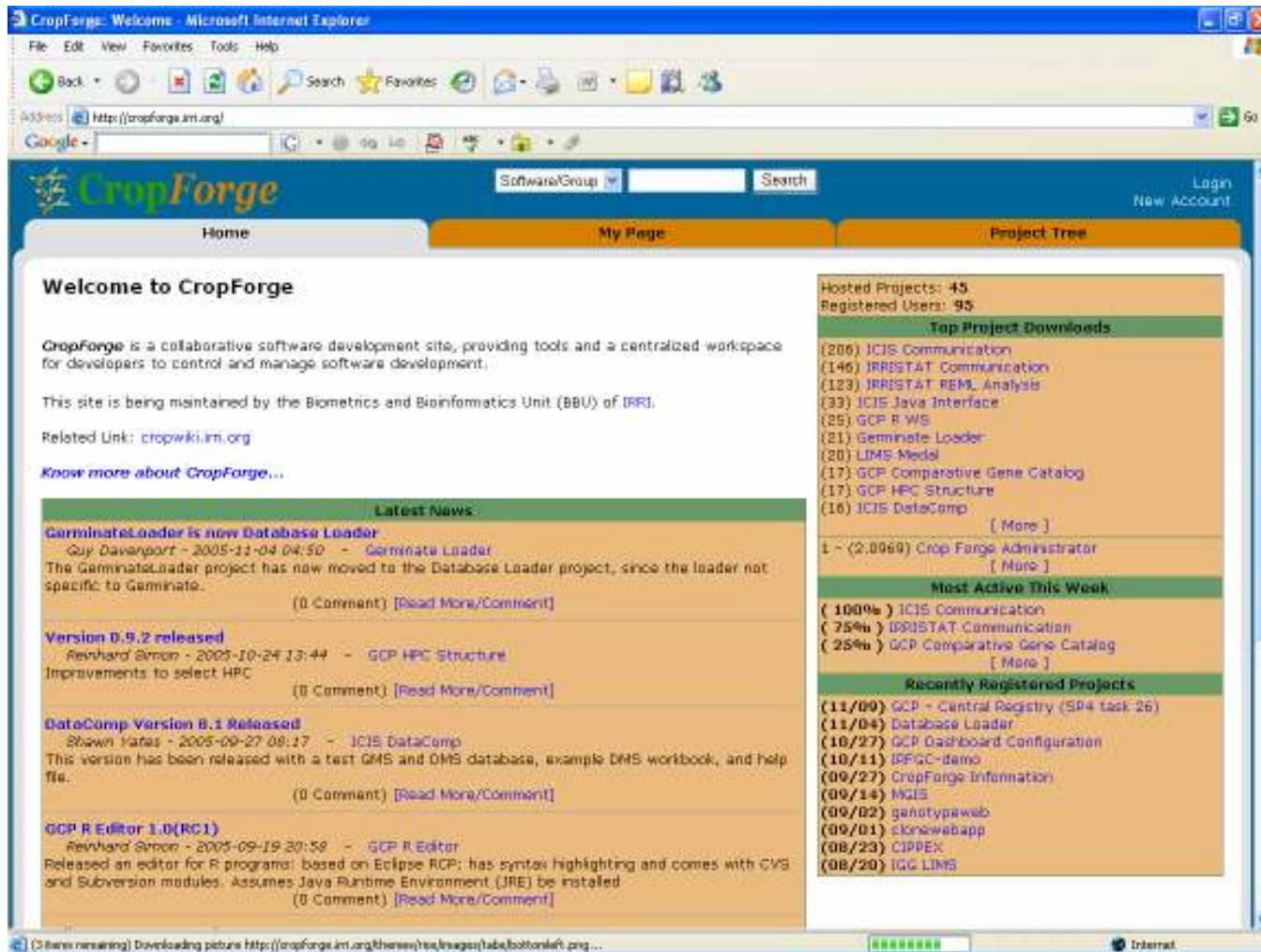




## 3<sup>rd</sup> Party Tools for GCP Integration

Tool	Purpose
SoapLab2	Remote computational services access
Taverna	Bioinformatics work flow management
Apollo	Genome sequence browser
Cytoscape	Visualization of networks
ATV	Phylogenetic tree visualization
JaView	Comparative sequence alignments
TMEV	Microarray data analysis
EASE, Mapman	Gene functional annotation
CMTV	Comparative mapping and QTL
MAXDLoad & MAXDView	Microarray data management
GDPC tools (Browser, Tassel)	Genomic diversity analysis

# GCP “Pantheon” Project in CropForge





## Closing Perspective

- The GCP is a global consortium of 22++ crop research partners who need to share diverse large data sets and tools, in a globally distributed manner.
- Given the scope and duration of the GCP, developers within the consortium embraced the task of developing public global informatics standards for interoperability and integration.
- The effort is an open source, global community building exercise.
- We welcome the participation of any and all interested scientists and developers who might wish to use and/or contribute to the further evolution and application of these standards.



# The Generation Challenge Programme Crop Informatics team and Contributors

*Theo van Hintum (WUR), GCP Subprogramme 4 Leader*



**IRRI-CIMMYT  
Crop Research  
Informatics Laboratory**



Graham McLaren

Thomas Metz

Martin Senger

Ramil Mauleon

Mylah Anacleto

Michael Jonathan Mendoza

Victor Jun Ulat

Arlet Portugal

Ryan Alamban

Lord Hendrix Barboza

Jeffrey Detras

Kevin Manansala

Jeffrey Morales

Barry Peralta

Rowena Valerio

Nelzo Ereful



Guy Davenport

Trushar Shah

Kyle Braak

Sebastian Ritter

Yi Zhang

Sergio Gregorio

Joseph Hermocilla

Michael Echavez

Roque Almodiel



Samart Wanchana

Supat Thongjuea



**CIRAD:**

Manuel Ruiz

Alexis Dereeper

Matthieu Conte

Brigitte Courtois

**Bioversity:**

Mathieu Rouard

Tom Hazekamp

Milko Skofic

Raj Sood

**NIAS:**

Masaru Takeya

Koji Doi

Kouji Satoh

Shoshi Kikuchi

**EMBRAPA:**

Marcos Costa

Natalia Martins

Georgios Pappas

**CIP:**



Reinhard Simon

Edwin Rojas



Jayashree Balaji

**ICARDA:**

Akinnola Akintunde

**NCGR:**

Andrew Farmer

Gary Schiltz

**SCRI:**

Jennifer Lee

David Marshall

**Cornell University:**

Terry Casstevens

Pankaj Jaiswal

Dave Matthews

**ACGT:**

Ayton Meintjes

Jane Morris

**University of British  
Columbia:**

Mark Wilkinson

**GSC Bioinformatics  
Graduate  
Program, BC  
Cancer Agency:**

Benjamin Good

James Wagner

