

Pretoria Platform Development

GCP Platform Developers Workshop - Mar 20-23, 2006

Contents

- 1 Monday, 20th March
 - 1.1 Standalone CMTV and implementation of spanning use-cases
 - 1.1.1 Presentation/Demonstration pdf
 - 1.1.2 Discussion
 - 1.1.3 ToDo until end of May 2006
 - 1.2 Demonstration of data retrieval using the Germplasm Domain Model
 - 1.2.1 Presentation/Demonstration
 - 1.2.2 Discussion
 - 1.2.3 ToDo until end of May 2006
 - 1.2.4 Postmortem
 - 1.3 Demonstration and discussion on use of CropForge
 - 1.3.1 Presentation/Demonstration pdf
 - 1.3.2 Discussion
 - 1.3.3 ToDo until end of May 2006
- 2 Tuesday, 21st March
 - 2.1 Demonstration of data retrieval using the Passport Domain model
 - 2.1.1 Presentation/Demonstration pdf
 - 2.1.2 Discussion
 - 2.1.3 ToDo until end of May 2006
 - 2.2 DIVA-GIS coupled to the Location model and ICIS5
 - 2.2.1 Presentation/Demonstration ppt
 - 2.2.2 Discussion
 - 2.2.3 ToDo until end of May 2006
 - 2.3 Demonstration of generic web interface and map viewers using the middleware mapping module

- 2.3.1 Presentation/Demonstration
 - 2.3.2 Discussion
 - 2.3.3 ToDo until end of May 2006
- 2.4 Mondrian data warehouse accessing passport, phenotyping and genotyping data
 - 2.4.1 Presentation/Demonstration pdf
 - 2.4.2 Discussion
 - 2.4.3 ToDo until end of May 2006
- 3 Wednesday, 22nd March
 - 3.1 Demonstrate web-based, OS independent sample tracking facilities
 - 3.1.1 Presentation/Demonstration pdf
 - 3.1.2 Discussion
 - 3.1.3 ToDo until end of May 2006
 - 3.2 Demonstrate IGMLIMS and IGGEMS in ASP .NET and Java implementations
 - 3.2.1 Presentation/Demonstration
 - 3.2.2 Discussion
 - 3.2.3 ToDo until end of May 2006
 - 3.3 Demonstrate platform integration of APOLLO, MAXD and Genoma applications
 - 3.3.1 Presentation/Demonstration
 - 3.3.2 Discussion
 - 3.3.3 ToDo until end of May 2006
 - 3.4 Demonstrate Moses Dashboard and MOBY implementation for a standard use case
 - 3.4.1 Presentation/Demonstration
 - 3.4.2 Introduction
 - 3.4.3 Biomoby registries

- 3.4.4 Documenting services
- 3.4.5 Robust.. prototyping is 2005, 2006 for production
- 3.4.6 GCP MOBY Environment - publishes known GCP services
- 3.4.7 AEAP - Support Services (Making Moby as easy as possible)
- 3.4.8 Dashboard Demonstration
 - 3.4.8.1 Dashboard simple client views
 - 3.4.8.2 Development/Registration of Web Services
 - 3.4.8.3 MoSeS
 - 3.4.8.4 Summary
- 3.4.9 ToDo until end of May 2006
- 3.5 Discussion (Short)
- 3.6 Summary and discussion of tasks needed to complete/refine 2005 deliverables
 - 3.6.1 Discussion
 - 3.6.2 ToDo until end of May 2006
- 4 Thursday, 23rd March - Host Presentation and Plans for 2006
 - 4.1 The ACTG Bioinformatics and Computational Biology Unit
- 5 Presentation and discussion of 2006 subtasks and deliverables
 - 5.1 GCP workbench middleware: framework and components(2006-16-1)
 - 5.2 Generalized query engine and result integrator (2006-16-2)
 - 5.3 Germplasm genotype visualisation tools (2006-16-3)
 - 5.4 Hibernate adaptors for the ICRIS database (2006-16-4)
 - 5.5 Genomedium interface for genotype, QTL and map data (2006-16-5)
 - 5.6 Genomic sequence analysis and visualisation (2006-16-6)
 - 5.7 Geographical data analysis and visualisation (2006-16-7)
 - 5.8 Web Service/Internet Data Source Integration (2006-16-8)
 - 5.9 Functional Genomics Tool Integration into Pantheon (2006-16-9)

Monday, 20th March

Standalone CMTV and implementation of spanning use-cases

- Presenter: Kyle Braak
- Rapporteur: Thomas Metz

Presentation/Demonstration

Overview of the Pantheon (platform development) Project with 4 sub-projects:

- Demeter - domain model specification
- Ceres - domain model implementation
- Belenus - application support
- Osiris - data source support

Overview of Eclipse plug-in architecture and development

Overview Eclipse Modeling Framework (EMF)

Overview of ISYS and CMTV

- CMTV relies on ISYS to connect to data sources and other tools
- ISYS will be Eclipse plugin

Integration of CMTV into GCP platform

- use cases
 - Query database(s) for available genomic maps and display maps individually or together
 - Query database(s) for available QTL distributions by various environmental conditions and display together
- Implementation plan
 - database of CIMMYT Map and QTL data (PostgreSQL)

- map the database onto the domain model using Hibernate
- provide API for querying model
- map required objects in the domain model into the ISYS model
- develop an Eclipse extension point for ISYS service providers
- Screen shots of prototype (prototype itself not working yet)
 - retrieve list of maps via ISYS

Discussion

- Eclipse almost in general GCP SP4 use. Should be a recommendation but not a requirement.
- Should have ways to start applications outside the Eclipse framework.
- Extension points are for communication between plug-ins, not only RCP plug-ins.
- GCP middleware is currently not designed to support application communication, ISYS may be needed for that.
- Requirement of use of GCP domain model and web services protocol. Freedom of data sources and data connectivity.
- Agreement was to use web services based on BioMoby. Biocase is not a web service, not enough to expose data to the GCP model.
- De-emphasised domain modelling 6 months ago, started with prototyping.
- Extending of Ceres for every database? Osiris yes, but Ceres should be database independent.
- Need for Adapter Classes in Osiris (Richard explain!)
- Reasons for implementing Eclipse ISYS in RCP need to be elaborated
- The added value of this work needs to be communicated (e.g. existing CMTV can now plug into the GCP domain model, ISYS Eclipse plugin), needs to be proven by installation elsewhere (CIRAD, ICRISAT)

ToDo until end of May 2006

- provide a download of a working prototype, fully featured CMTV

- develop an update manager for ISYS Eclipse, that will automatically look for new ISYS components as they come in
- documentation of choice of RCP
- documentation on implementing data sources in the platform using Hibernate
- redefine user documentation for GCP-integrated CMTV

Demonstration of data retrieval using the Germplasm Domain Model

- Presenter: Richard Bruskiewich
- Rapporteur: Akin Akintunde

Presentation/Demonstration

- The objective of this platform subproject is to implement the germplasm/phenotype/genotype component of the GCP domain model (see <http://www.generationcp.org/model>).
- The use case driving development of this component is the Pantheon Querybuilder, whose initial use cases are given here.
- Querybuilder code is on CropForge the query builder is on CropForge CVS under the here.
- Sample demo of the Pantheon Querybuilder with ICIS as a back end data source for germplasm information was attempted. Unfortunately because of one very nasty Eclipse RCP bug in the program which he could not resolve prior to his presentation time (which he described down to single-stepped Java source code level), he could not fully demonstrate the software in a single, vertically integrated manner. Instead, three distinct software runs, as follows:

An earlier partial implementation of the querybuilder highlighted the management of controlled vocabulary and ontology terms in the querybuilder (fully working prototype).

The successful retrieval of IRIS germplasm identifiers based on a name search, using the back end Osiris plugin and JUnit tests, was successfully demonstrated

A "mock data" run of the graphical user interface (GUI) of querybuilder, showing the use of the class navigation tree, the general search box, the result synopsis view and the result list views.

- An Eclipse plugin configuration issue currently (as of demo) connection of the back and front end components of querybuilder (intensive troubleshooting underway at this time).
- The presentation included a general discussion of the use of the Spring (<http://www.springframework.org>) lightweight enterprise bean framework. Books on the subject from <http://www.manning.com> are recommended reading:
 - **POJOs in Action:** by Chris Richardson outlines a domain model driven, test driven approach using Spring plus Hibernate (or JDO) designs
 - **Spring in Action:** by Craig Walls and Ryan Breidenbach is a detailed reference of Spring
 - **Hibernate in Action:** by Christian Bauer and Gavin King is also useful.

Discussion

- Q: Why Spring? How does this complement or diverge from the current (CIMMYT++) Session object strategy?
 - Ans: Spring allows for "Dependency Injection" (DI) that cleanly decouples class implementations from one another, allowing for rewiring the code without changing/breaking it so readily. Also permits effective mock testing of class singletons (hard to do with static variable approach). Might help us clean up the wiring up of Osiris to Ceres (make it less "hard coded")
 - Counterpoint: Jody of uDig says Eclipse plug-in extension mechanism is a somewhat opposite philosophy to Spring DI. We probably need to consider the implications of this assertion very soon.

ToDo until end of May 2006

Complete a working implementation of basic QueryBuilder functionality

Fix the Elipse plugin bugs

Add full Entity class browsing and querying directly generated from the domain model

Add detail text and some basic specialist graphical views (e.g. Germplasm pedigree, Study table data, etc.)

Couple map/marker data with CMTV visualization(collaborate with

CIMMYT and CIRAD)

Couple location data with DIVA visualization

Full integration of ICIS as back end source for
germplasm/phenotype/genotype/passport/marker/genotype data from ICIS

Developer documentation of Querybuilder/Pantheon (Belenus/Demeter/Ceres/Osiris)
layers

User documentation and tutorial for Querybuilder

Postmortem

- Upon discussion (Tuesday) with Jody from uDig, and some internet searching, there seems to be an probably source of the "bug from hell" whose resolution is believed to be documented at <http://www.hibernate.org/311.html>. Unfortunately, application of this fix is pending a fuller understanding on how to apply it.

Demonstration and discussion on use of CropForge

- Presenter: Thomas Metz
- Rapporteur: Graham McLaren

Presentation/Demonstration

Discussion

ToDo until end of May 2006

- Better support for subversion
- GCP communication project vs. software development projects, try to make a case (Graham, Richard, Thomas)

Tuesday, 21st March

Demonstration of data retrieval using the Passport Domain model

- Presenter: Mathieu Rouard
- Rapporteur: Richard Bruskiewich

Presentation/Demonstration

- Multicrop passport data
- Use cases documented in UML: Filter on passport data
- Phase I data model - somewhat MCPD Entity-Relational (relational database) specified
- Phase II data model (current EclipseUML fully compliant with overall GCP domain model for germplasm, etc.)
- Passport data model prototype
 - ICRIS data source @ ICRISAT
 - Hibernate mapping (Osiris)
 - domain model layer (Demeter/Ceres)
 - Web service layer (Belenus)
 - Tomcat + Axis
 - Client Taverna
- ICRIS passport data tables (existing)
 - Most information in one table (Passport) with a few support tables (relatively simple)
- Persistence implementation
 - Hibernate mapping & properties
 - Configuration/SessionFactory/Session

- Implementation
 - Elaborated a Demeter class
 - Implemented a Ceres class
 - MOBY Services Belenus layer
 - Need for MOBY data types
 - Need for service semantics
- Richard, Martin, Jayashree and Mathieu: iterated on MOBY data type specification from Domain model
 - Not fully compliant to GCP domain model but practical for prototype
 - Registered in MOBY Central @ GCP (IRRI)
- Web services:
 - Get an accession list based on the taxonomy
 - Get an accession list based on the collection name
 - Get an accession list based on additional descriptors
 - Get passport details for an accession id
 - Get passport details for an accession id set

RB note: could 4 & 5 be one web services (#4 is just a input set of "one"?)
- Used MOBY dashboard to design/deploy/test MOBY services
- Taverna: MOBY client (Scufl workbench)
 - Designed Taverna work flows...
- Musa/MGIS Java data entry tool demonstrated which uses Hibernate mappings onto a MySQL database
 - Question: could this tool be generalized to other crops and databases and to have a GCP domain model connection

Discussion

- Use cases and interfaces to be elaborated
 - GCP domain model may diverge from conventional genetic resources

concepts; need to ensure seamless conversion so as not to disrupt end user and genetic resources manager's management of the world

ToDo until end of May 2006

- Finalize prototype by May
- Fully document prototype and "lessons learned"
 - Structured by audiences: users, admin (web services deployment) and developer
- Extend the prototype to other crop/site passport data (wrapped using Hibernate)
 - CIP and IRRI (IRIS/ICIS) Osiris connection

DIVA-GIS coupled to the Location model and ICIS5

Presentation/Demonstration

- Presenter: Edwin Rojas
- Rapporteur: Kyle Braak

Discussion

Questions?

- Can we leverage any open standards to make our job easier - This is the question
 - Which specific one (os)?
 - Currently, we are shopping for standards.
 - GIS is well established in regard to OS, but the community has to drive which standard is incorporated.
- Can DIVA connect to distributed data sources via Pantheon?
 - "Optimistic feelings expressed"

Deliverables?

- Within the time frame of the project (over the next 2 months until May 2006), can we integrate DIVA into Pantheon?
 - WholeHearted YES
- From a coders perspective, GIS queries should be achievable via a DIVA plug-in.
- Query builder is perhaps the most appropriate tool to be able to integrate DIVA with Pantheon.
 - Ideally, a user will see a DIVA pop up menu (provided via DIVA plug-in) and essentially be able to have DIVA at their fingertips.

ToDo until end of May 2006

- Ultimately, we would like to create a DIVA plug-in for Query builder opening up the user to its inherent GIS capabilities
- By this deadline, we need to integrate DIVA into Pantheon. How this is achievable, whether it be via the above-mentioned, or within DIVA itself is to be determined.

Demonstration of generic web interface and map viewers using the middleware mapping module

- Presenters: Manuel Ruiz and Alexis Dereeper
- Rapporteur: Balaji Jayashree

Presentation/Demonstration

Highlights:

- Implementation of the core of the mapping middleware module (Demeter- Ceres)
- Database adaptors : Hibernate + Web Services (Osiris)
- Generic Web Application (Belenus)
- Connection to existing generic Map Viewers (Belenus)

The module prototype was developed in Java.

Database Adaptors were generated for two different types of databases:

- Generic databases: CMAP, GBrowse and Chado(IRRI)
- Specific databases: Tropgene(CIRAD via BioMoby), EMBL(via EBI web services),Gramene(via hibernate)
- generic viewers connected: CMTV, CMAP and Gbrowse
- The mapping data model was presented (available on the wiki)
- Users using generic databases may use the GCP web pantheon, while those using specific databases need to develop wrappers.

Demonstration of the pantheon web services:

- Genotype/mapping domain model
- Datasource specific implementation
- Platform architecture
- Demonstration of the use of the application

The application servlet along with its dependencies(Java, MySQL DB, Tomcat,and pantheon plugins) was provided for installation and testing.

Test the demo :

<http://sat.cirad.fr/sat/GCP/servlet/org.generationcp.application.mapping.servlet.servlet>

Discussion

- A good demonstration of the platform using varied data sources(generic and specific).
- New use cases development in progress.
- The application qualifies for a demonstration at the GCP annual meeting, as a useful deliverable for the user.
- Updating to the latest version of the model was thought not to be difficult.
- Every institute using CMAP data would find this application useful

ToDo until end of May 2006

- Documentation of the program
- Get Feedbacks from GCP users

- Development of required use cases by users

Mondrian data warehouse accessing passport, phenotyping and genotyping data

- Presenter: Edwin Rojas
- Rapporteur: Martin Senger

Presentation/Demonstration

- Data Warehouse motivation and 2005 – 2006 activities
- Data Warehouse Architectural
- Data Warehouse Types – Part I
- Data Warehouse Types – Part II
- Multidimensional Model Elements – Part I
- Multidimensional Model Elements – Part II
- Mondrian engine versions
- Open Source Components for Research Intelligence(RI) Platform
- Data Warehouse Technology for GCP
- Relationship of GCP Domain Model to Platform
- Relationship of GCP Mondrian Data Warehouse to Platform
- Case Study for ICIS Inventory (IMS) Database
- Case Study for CIP Passport Database
- Case Study for CIP Field & Molecular Experiments Database
- Case Study for ICIS Genealogy (GMS) Database
- Who is testing GCP Mondrian Data Warehouse ?
- Demos and Tutorial
- Pending Activities and Tentatively Activities for 2006

- Important New Features on work, for next version

Discussion

ToDo until end of May 2006

Wednesday, 22nd March

Demonstrate web-based, OS independent sample tracking facilities

- Presenter: Balaji Jayashree
- Rapporteur: Mathieu Rouard

Presentation/Demonstration

Collaborative project between ICRISAT, ICARDA and CIMMYT

- Not part of the platform.
- data capture software for the genotyping workflow.
- a measure of quality for data generated data
- provides for an algorithm to auto-bin genotyper output data
- the software has four modules that allow starting an experiment, sample tracking, report generation and storage of details on protocols and markers.

Sample tracking relative to:

- DNA extraction
- DNA dilution
- PCR plate design
- Marker selection

- Electrophoresis
- upload of output files, merging and automated allele binning.
- Application being used at ICRISAT, tested at CIMMYT, being evaluated for transfer to IITA, ICRISAT-Nairobi and ILRI.

Discussion

- how does the system allow for tracing back of samples using a LIMSId. especially with situations like splitting of sample data/combining germplasm across studies/plates?
- User requests from IITA, ILRI and ICRISAT-Nairobi were discussed.
- Code to be seen by IRRI developers
- Can possibly feed into domain modeling task.

ToDo until end of May 2006

Distribution of the system at ICARDA and CYMMYT

Demonstrate IGMLIMS and IGGEMS in ASP .NET and Java implementations

- Presenter: Akin Akintunde
- Rapporteur: Manuel Ruiz

Presentation/Demonstration

- Demonstration of ICIS GEMS
 - Integration of 4 different modules : LIMS, Web, Storage, GEMS
 - Delphi for standalone part
 - .dot net for Web application
 - The system provides navigation capability

- Main services
 - GEMS information
 - allows PCR tracking
 - data can be exported
 - XML file with all data based on the GCP template
 - HTML, etc.
 - My Workspace : manage personal study informations
 - Get personal accessions information
 - DNA extraction
 - Plates information management
 - PCR Tracking
 - Molecular Variant Advanced Search
 - Web application
 - PCR Advanced Search
 - Web application
- Storage management (user, administrator)
 - Login
 - Filters
 - Items
 - Several options for display
- Setting and Options
- Presentation ppt
 - A set of the visualization tools are being planned for the different aspects of the GEMS
 - Comprehensive user manual to be done

Discussion

- ICARDA needs capacity for SP1 integration data : data curator (problem of staff management?) (same issue with ICRISAT)
- Curation is the job of those generating data, the data have to be delivered in the good format

ToDo until end of May 2006

- Program documentation
- To make a start on java equivalent (or database wrappers)
- Source code in Cropforge

Demonstrate platform integration of APOLLO, MAXD and Genoma applications

- Presenter: Richard Bruskiwich Masaru Takeya and Marcos Costa
- Rapporteur: Edwin Rojas

Presentation/Demonstration

- Recruited several University of Philipines students

Martin Senger assisted for SoapLab and MOBY services development for IRRI, EMBRAPA and NIAS Sequence inventory tool MaxD report by Masaru Genoma report by Marcos

Show sequence inventory tool for Web (<http://localhost/SequenceTemplate>)

- Add entry feature, alignment
- Search NCBI module

Masaru Presentation Show Three tier model of gene expression platform and database Show RED(Red Expression Database) functionality

Data Mining Tools Show Flowchart to find Cis-Elements

Show functional genomic web services

Marcos Presentation Brief demonstration of genomic system from Embrapa Show architecture design using proxy before firewall, using business delegate component, designed in Corba The Embrapa Genomic system code is free distribution

- Demonstrations:
 - SequenceInventory (Richard)
 - NIAS/RED/MAXD (Masaru)
 - Genoma (Marcus)

Discussion

RED1=Rice and RED2=Other crops Products to be available in 2 years and developing has started. Concentrated more in design the database Different piece of data needed to be connected How integrate Embrapa Genomyc System to GCP QueryBuilder, possibility to change Proxy to support Moby What is Embrapa contributions to GCP ? Public use case How is the documentation (Pending to finish internationalization) To install and test in Pretoria and IRRI

ToDo until end of May 2006

- UP students on “summer break” and have pledged to pay more attention to the project, hopefully leading to more/refined prototypes of tool integration
 - Sequence inventory tool refined/debugged by end user feedback from GCP postdoc
 - Chado wrapped further with GCP domain model using Hibernate and applied to tool integration task
 - Querybuilder (general searches on genomic data?)
 - GMOD tools: Apollo(?)
 - JalView(??)
 - Soaplab client for HPC connection integrated into platform?
 - MAXD integration with NIAS (maybe via ISYS?)
- Review documentation and translation for EMBRAPA System to be deployed at ILRI (Etienne De Villiers)

Demonstrate Moses Dashboard and MOBY implementation for a standard use case

- Presenter: Martin Senger
- Rapporteur: Richard Bruskiwich

Presentation/Demonstration

Side note: Subversion reference book free online: <http://svnbook.red-bean.com>.

Presentation (PDF) and a powerpoint tutorial.

Introduction

- GCP biomoby infrastructure
 - (mostly) not about individual services
- .. not realtime now (here)
- IRRI fingerprints all over the work w/ GCP colleagues
- Important:
 - to have services, but...
 - must be robust technology to support good services
- Big picture: manages semantics via standard protocol
 - Central, Services, Clients
- Bottom line:
 - service providers are responsible for services
 - community is responsible for biomoby community data types
 - Register your data types

Biomoby registries

one (default) in MOBY Central Canada

only post when ready, deregistration policy is coming...

now: can't protect against hackers: everybody can now remove service

future (2-3 months): there will be an RDF service definition will contain RDF endpoint path to a RDF document

if the RDF document is gone, service is removed (after several attempts...)

IRRI biomoby registry

for testing and development

Canada->Philippines mirror

code in CropForge

Documenting services

- <http://moby.generationcp.org> static information page
- <https://cropforge.org/projects/gcpmoby>
- email list service

Robust.. prototyping is 2005, 2006 for production

- Solve error conditions NOW
- Write test units NOW
- Let other people test your services NOW
- Document the code NOW
- Keep list of changes and TODOs
- S&S ("Show And Share" your code)

GCP MOBY Environment - publishes known GCP services

- moby.generationcp.org/mobyenv/checkservices
- Check out code for MOBY environment from Cropforge and use
 - Docs on web site as well as checkout docs
 - Cronjob mechanism can be used for locally failed services...

AEAP - Support Services (Making Moby as easy as possible)

- MOSES completed last year
- Dashboard is new and can now run services
- Registry Viewer Application which reads information for the IRRI registry and displays what the registered entry will do. Datatypes and services.

Dashboard Demonstration

- fire up at command line (jMOBY)
- Has panels for specialist purposes:
 - Registry browser
 - Register services
 - MoSeS code generation
 - Simple client
 - DebuggingPanel
 - Button to garbage collection

Dashboard simple client views

- Specialist views added by Java SPI (older alternative to Eclipse extensions?)
 - Image, pedigree viewer
 - Dashboard errors: provides workarounds of getting around development problems (like incorrect MOBY central service registration)

- can browse services (various ways), data types, service types

Development/Registration of Web Services

- Registration Panel
 - Data type registration
 - Service registration
 - Namespace registration
 - Service type registration

MoSeS

- Steps:
 - design service, autogenerate code & documentation (JavaDocs)
 - NEED TO WRITE BUSINESS LOGIC TO SUPPORT (MoSeS just mediates the transition between business logic and MOBY)
 - make a Jar file of the service implementation (Ant jar)
 - deploy to TomCat (sometimes tedious and problematic: check that all files transferred)
 - fill in services deployment info
 - e.g. where is Tomcat?
 - what is mapping of service name to service implementation?
 - Class name needs to be upper case (first letter)!
 - Deploy button (to deploy) - Tomcat must be alive!
 - Reload "axis" via Tomcat manager

Summary

- BTW, Dashboard uses "Blackboard" pattern
- easy to add panels, viewers, look and feel, help(?)

- documentation available in jMOBY or at ...

ToDo until end of May 2006

- Generated Moby data types from Demeter
- Mosts in Perl
- BioCASE (Tapir) -> Perl
- More panels/functions in Dashboard
- Maintenance
- Support for service providers

Discussion (Short)

- Question/concern: some worried about about technical longevity of MOBY
- Responses:
 - MOBY-S Canada is funded again (3-4 years) for some MOBY-S maintenance
 - S-MOBY (USA) is funded under NSF for 3 years (Virtual Plant Information Network)
 - Core technology is industry standard... MOBY adds bioinformatics semantics... GCP is developing domain model semantics.
 - We can host our own MOBY Central (already do...)

Summary and discussion of tasks needed to complete/refine 2005 deliverables

- Presenter: Graham McLaren
- Rapporteur: Thomas Metz

Discussion

- reviewed all ToDo lists
- reporting for of the entire project
- all source code on CropForge

ToDo until end of May 2006

Thursday, 23rd March - Host Presentation and Plans for 2006

The ACTG Bioinformatics and Computational Biology Unit

Presentation by Fourie Joubert

Presentation and discussion of 2006 subtasks and deliverables

GCP workbench middleware: framework and components(2006-16-1)

Martin Senger

Proposal 2006

Generalized query engine and result integrator (2006- 16-2)

Richard Bruskiwich

Generalized query engine and result integrator □ Richard Bruskiewich (IRRI). A generalized query engine and user interface (based on Eclipse RCP and code named □Pantheon□) was conceptualized in 2005. The functionality of this tool will be reminiscent of ACEDB (<http://www.acedb.org>). This tool will query by identification and by features the diverse entities of the GCP domain model to create sets of results, then permit invocation of specialist tools/views on individual specialist entities, as applicable (e.g. CMTV for map data, Apollo for sequence data, DIVA for location data, etc.) The GCP Querybuilder project has been created in CropForge for the development of this tool). The tool will also be connected to ICIS, Chado, ICRIS (ICRISAT) and other pertinent database schema adapters. ICRISAT (Jayashree Balaji) will collaborate with IRRI on germplasm and phenotype visualization interfaces for this tool, and data source adapter development for ICRIS.

Proposal 2006

Germplasm genotype visualisation tools (2006-16-3)

Akin Akintunde

Presentation

Germplasm Genotype Visualisation □ Akinnola Akintunde (ICARDA). This new application is projected to have the following features:

- A 3D Allele viewer: to present the molecular variant information of different plants and

primers in a 3D image form, providing visual stimuli to researchers in addition to the usual numerical displays.

- A tree□like pedigree display from the Molecular output display frames.
- An Integrated allele mining related display
- A Micro Satellite Finder (finds the DNA Micro Satellite in Sequences)
- The tool will also be connected to the ICIS database schema adapters.

Proposal 2006

Hibernate adaptors for the ICRIS database (2006-16-4)

Balaji Jayashree

Presentation Our activity under this project is to develop adaptors for passport, genotyping data in ICRIS. This should be possible to do once the revised domain models, use-case

and Moby data types become available. A prototype for passport / genotyping data should be possible to do by end-August provided we can reasonably agree on the data types and begin to work on it now. Martin has assured us that he will return to us with news on the updated domain models and use-cases. We should be able to get started once this is available.

Proposal 2006

Genomedium interface for genotype, QTL and map data (2006-16-5)

Kyle Braak

Presentation

QTL and map data analysis and visualisation □ Guy Davenport (CIMMYT)

- The pilot project to integrate the NCGR Comparative Mapping and Trait Viewer

(<http://www.ncgr.org/cmtv>) will be completed in 2006.

- The mapping domain model will be improved by integration of links between QTL

mapping and sequence data to facilitate comparative mapping.

Proposal 2006

Genomic sequence analysis and visualisation (2006-16-6)

Manuel Ruiz (CIRAD)

Presentation

Genomic Sequence analysis and visualisation:

- The Java□based GMOD “Apollo” genome browser will be integrated into the platform.
- A generic Eclipse RCP interface for mapping data queries to Osiris interfaces to

databases of interest: Gramene, Chado, TropGene

- **Generic adaptors:** using BioMoby web services, between Perl GMOD interfaces

(GBrowse, CMap) and the GCP Middleware.

Proposal 2006

Geographical data analysis and visualisation (2006-16-7)

(Reinhard Simon), Edwin Rojas

Presentation

- Geographical data analysis and visualisation □ Reinhard Simon (CIP)
 - A pilot project to integrate the DIVA GIS tool in 2005 (<http://www.cipotato.org/diva>)

will be completed in 2006.

Proposal 2006

Web Service/Internet Data Source Integration (2006-16-8)

Martin Senger

Presentation

Web Service/Internet Data Source Integration □ Martin Senger (IRRI). Several priority internet data exchange protocols and tools will be fully embedded into the GCP platform, in collaboration with the GCP commissioned research task for web service application Development:

- **Autogeneration of MOBY data types** by Eclipse Modeling Framework technology(?)
- **BioMOBY**: extend MOSeS tool kit to Perl
- **BioCASE (wrapped as BioMOBY services)**
- **SoapLab2**: next generation of SoapLab (developed and deployed to HPC)
- **Genomic Diversity and Phenotype Connector (GDPC;** <http://www.maizegenetics.net/gdpc>) to enable access to GDPC wrapped data sources and to permit connection of GDPC compliant tools like Tassel and the

GDPC browser.

- **Taverna Workflow Engine:** will be connected to the GCP domain model to allow the integration of rich web service work flows in the platform.
- **Enhancement of tutorial and training resources:** for all of the above
- Coordinate GCP integration with the Virtual Plant Information Network
- Provide technical support for IRFGC MOBY Rice Network deployment.

Proposal 2006

Functional Genomics Tool Integration into Pantheon (2006-16-9)

Richard Bruskiwich, Marcos Costa, Masaru Takeya

Presentation

Gene expression data analysis and visualisation □ Masaru Takeya (NIAS). The Java□based MAXD gene expression software and database system (<http://www.bioinf.man.ac.uk/microarray/maxd/>) will be customized for the GCP platform, in collaboration with the 2005 task (32) for “Gene Expression Repository and Data Mining Tools”.

Proposal 2006