

GEMS development & strategy

ICIS 2007 Workshop

T. Metz, IRRI

Genotyping Data

- Genotyping data are stored as studies in DMS
- Genotyping data are loaded with (extended) workbook
- Scripts for loading data are also available
- Practical experience with SSR and DArT genotyping data (+isozyme)
- Genotyping data are outputs/primary data

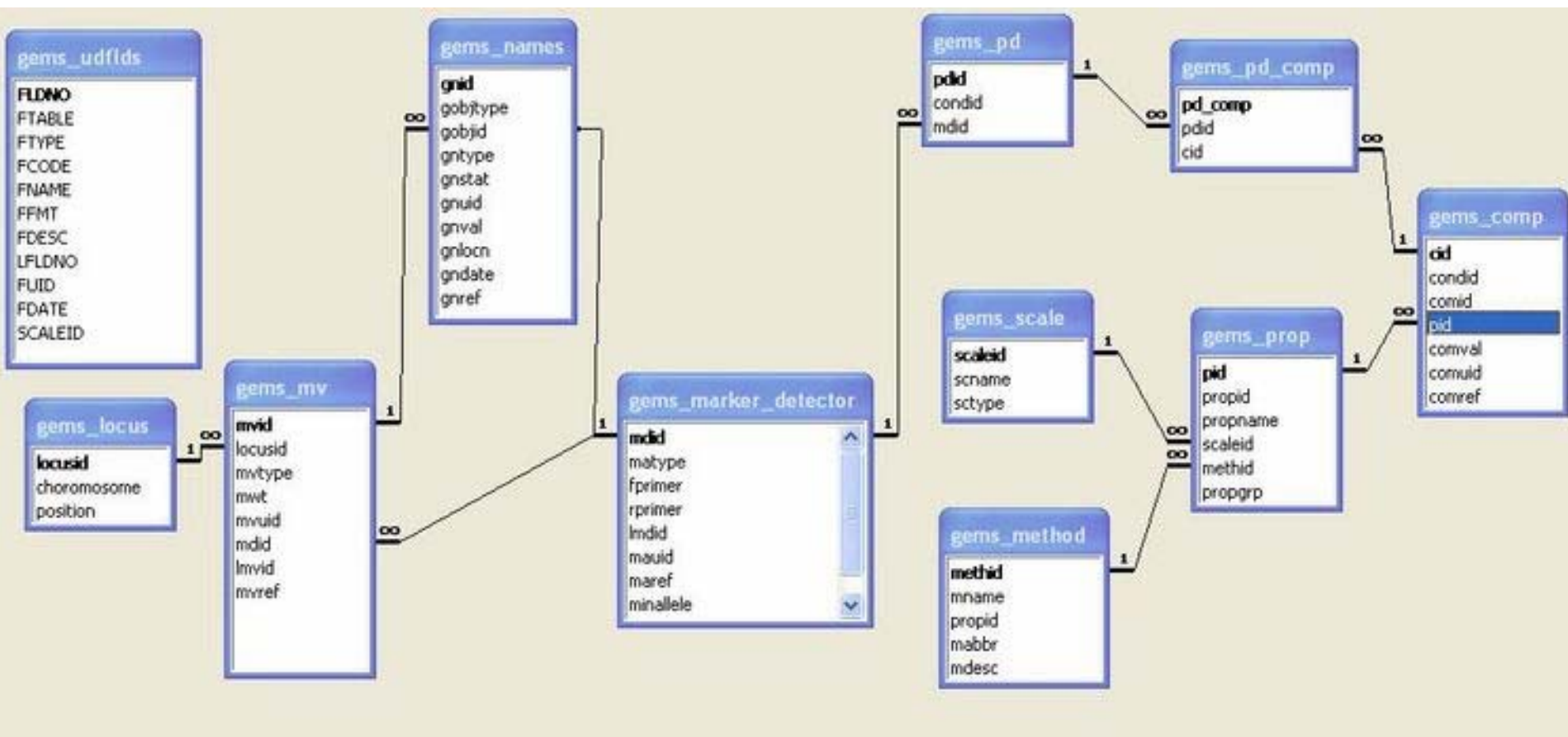
Marker Data

- Marker data are stored in ICIS GEMS
- Marker data are loaded with workbook (ID's, names, protocols)
- Standalone (prototype) interface for marker data exists (SSR)
- No complete sets of marker data are available in GEMS
- Marker data are often inputs/secondary data in genotyping studies

GEMS

- Schema
- Data loading tools for genotyping and marker data (workbook & scripts)
- Data curation tool for marker data

GEMS schema



GEMS - Quo vadis?

- Scope
 - Objects: alleles, markers, protocols, loci, maps, ...
 - Coverage: all published data for a crop?
 - Details: e.g. protocols
 - Technologies: SSR, DArT, ...

GEMS - LIMS

- Sample tracking
 - GMS (GIDs & germplasm methods)
- Inventory management
 - InTrack
- Protocols
 - GEMS
- Full LIMS
 - Workflow automation, process control, instrument interfaces, standards, controls, chemicals management, laboratory supplies

GEMS – ICIS32 DLL

- 20+ DLL functions for GEMS have been implemented
- Mainly related to marker data loading via workbook

GEMS – GCP ICIS datasource

- First use cases for GEMS data retrieval as part of GCP ICIS datasource have been implemented
 - Get list of markers
 - Get list of alleles

GEMS – input & curation

- Batch loading with workbook
 - Genotyping and marker data
- Standalone tool for marker data curation
 - Current MS Access prototype (SSR)
 - Re-implement in Delphi with separate interface for each marker technology?
 - Re-implement as a web-based interface?

Data retrieval – Wheat CRC

- Select, export and download genotyping datasets for use in other tools
 - Genotype visualisation tool (SCRI)
 - Graphical genotyping tool (WUR)
 - Heatmaps (R-statistics)
- Crosstab queries for phenotyping data
 - Germplasm x study
 - Trait x study
 - Overlap of germplasm in studies over time

Data retrieval – Wheat CRC

- List-driven retrieval
 - Germplasm list + marker list => Genotyping data
 - Germplasm list + trait list + study list => Phenotyping data
- Co-occurrence
 - Germplasm evaluated in pairs of studies (by single trait or set of traits)

Merging phenotyping and genotyping data in a query

- Many-to-many link
 - Multiple observations for a germplasm x trait combination (data analysis problem)
 - Repeat genotyping for a germplasm x marker combination (consensus genotype problem)
- One-to-one link
 - Studies with mean phenotype results per germplasm x trait and consensus germplasm genotypes can be merged

Data Warehouse

- Stored procedures in PostgreSQL and MySQL to pre-assemble phenotyping and genotyping data