Diversity Arrays presents KDDart, a new IT infrastructure we developed in conjunction with Australian and international partners and collaborator organisations. The KDDart platform, which is in early stages of deployment and testing within several projects, is a modular platform designed to integrate high data volumes that often originate from various sources.

The core module, which collects field data for plant performance, can easily be ported to collect any kind of phenotypic data from both breeding and natural plant and animal populations. The environmental module can store from high to lower resolution spatio-temporal data, obtained through sensor devices. These devices can be configured to communicate with KDDart remotely or from publicly available GIS/remote sensing data sources. Both phenotypic and environmental modules are connected to molecular data module enabled to store marker data from high-volume, high-resolution genome profiling technologies (e.g. DA'Tseq). In addition, the system is capable of storing inventory and pedigree-related information.

KDDart is designed as a modern IT infrastructure with 3 major components (layers):
1. Database layer
2. Application layer RESTful API (Data Access Layer or DAL)
3. Client Applications/Software layer

The Application layer API is a heart of the system and is designed as a RESTful web service which provides and manages access to the KDDart database. DAL is primarily a set of generic methods to perform all possible database operations and is exposed as a web service RESTful API.

Client Applications layer is a collection of scripts, software apples etc. accessing the database through the DAL. These applications can be written as web, desktop or mobile applications using any modern programming language.

Data collected and stored in KDDart is available for statistical analysis or data mining using KDCompute, which represents KDDart's generic plug-in for batch processing. Processed data/information can be viewed, further analysed and interpreted by project staff and other system users through KDDart's user secure and controlled environment.

The KDDart system provides its clients with a platform to integrate our marker services with customers' phenotypic and environmental data, enabling more efficient data management and analysis. The scenario illustrated below describes how the genotyping data can be connected to the KDDart platform and analysed together with all other data types on clients' platform. This integrated system configuration, facilitated by the web service (RESTful API) architecture of both systems, provides clients a more compact system resulting in a better user experience.

To learn more about DART's genotyping service, please see the poster: "DA'T™ and DA'Tseq™ Genome Profiling for Breeding, Pre-Breeding and Population Genomics" with number P0052.

KDDart Platform for Data Integration in Pre-Breeding, Breeding and Ecology

**KDDart: Knowledge Discovery in Database art**

**KDDart Knowledge Discovery in Database art**

**KDDart Features**
- Integrating all data types relevant to breeding, crop management and/or ecology (genetics, field data and environment)
- Remote sensing and high density environmental monitoring devices (remote sensors can access the system directly)
- Seamless integration of analytical /data mining tools (e.g. through our own KDCompute plugin infrastructure, but also as stand alone tools)
- Immediate integration of modelling technologies with KDDart
- Enabling informed decisions to breeders, farmers and research community
- Creating an "environment" for system growth and improvement through API (RESTful services)

**Data Structures**
- Phenotypic data is organised into experiments
- Experimental designs and traits are definable features of every implementation and may utilise existing ontologies
- Granular germplasm information with pedigree is connected to both environmental data and marker data
- Genotypic data and maps (both genetic and physical) can be saved in the system
- All experimental data can be spatially annotated and experiments organised into workflows
- All germplasm can be organised into inventory and all samples can be organised into storage locations
- Sensors and other devices can be used to enhance capture of phenotypic and environmental data and help in inventory management
- Environmental and other spatial data can be stored in GIS layered structures and related to other data through coordinates
- Spatio-temporal data can be stored in environmental modules
- Most data entities can be extended with custom defined additional data types and multimedia can be attached

**Applications**
- **KDSmart** - designed to operate on a variety of handheld devices for field data collection. Originally developed from a stable version of DataCapture software jointly developed by Q/NSW DPI, it has been adapted to meet a variety of new user requirements. Online mode allows direct synchronisation of trial data with the KDDart database. When off-line the trial data is available for the user to work in the field to capturing trial results.
- **KDMan** - the general system management utility. A web based application designed to setup system environment to satisfy entity dependencies, perform day-to-day data managerial tasks, manage and curate datasets.
- **KDXplore** - a versatile application, useful for breeders, technicians, curators and developers. Able to assist with your trial selection and manage the distribution of those copies onto multiple KDSmart devices. Data collection through the field with KDSmart can move back to KDXplore for data curation then upload, to ensure only quality trial data is stored in KDDart.
- **KDSens** - an application providing an interface between various generic environmental sensors, such as weather stations, soil probes, etc. and the KDDart database using the Data Access Layer (DAL). Sensor information is stored within KDDart.

**Your own application** - open architecture and Data Access Layer allow users to create their own scripts and software with custom interfaces and functionalities. They may range from simple ad hoc scripts up to large, complex software tools with entirely custom designed interfaces.