Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 209-212

HELICAL: A FULL-STACK AGILE PLATFORM ENABLING DATA FLYWHEELS AND DATA SHARING TO EMPOWER MODERN BREEDING PROGRAMS

D.P. Garrick[1,2,3], D.J. Garrick[1,2] and J.E. Howse[1]

[1] The Helical Company Limited, Rotorua, New Zealand
[2] Theta Solutions LLC, Lacey WA, USA
[3] A.L. Rae Centre for Genetics and Breeding, Massey University, New Zealand

SUMMARY

Helical is a comprehensive and fully integrated but modular informatics software-as-a-service (SaaS) platform that can be deployed as a complete breeding program information and decision support system. Built-in modules can standalone such as the genomic (SNP) database. Data import and extraction is highly flexible to cater for integration of any module with existing legacy systems. Licensees are typically research or breeding organisations who deliver and/or ingest data from stakeholders at an a) administrative level, b) user level, and/or c) a public level. The platform's architecture provides highly configurable but secure access controls for independent but integrated portals which are separately catered to each audience. At an organisation level, the system delivers full control of genomic data while maintaining freedom to work with multiple genotyping providers and service partners. This paper describes the platform's technical implementation and demonstrates its practical impact through commercial case studies, highlighting how organisations can continuously extract value from their genomic investment through sophisticated analytical tools. The platform's multi-tiered modular access facilitates powerful data sharing arrangements between stakeholders such as breed societies and commercial producers, while maintaining data sovereignty for each party. The result is a future-proofed system that enables organisations to maximise utility from their genomic data.

INTRODUCTION

Animal breeding programs face many complex challenges in today's technological landscape, including managing increasingly sophisticated genomic datasets, exchanging data with competing genomic labs and other service providers and linking a variety of data between stud and commercial operations. Advancement of DNA testing technologies has led to an explosion in genomic data volumes, yet most organisations remain constrained by old-fashioned inflexible systems that create data silos and lack the agility to incorporate new technologies, tackle emerging opportunities, or explore new service offerings without substantial manual intervention or costly investment.

The Helical platform addresses these challenges through a full-stack architecture that can be deployed either as a comprehensive information system or in a modular fashion such as a standalone genomic database that can be integrated with existing pedigree and performance systems. This flexible deployment model enables organisations to maintain part or all of their current registry systems while gaining sophisticated genomic capabilities, including the freedom to work with multiple genotyping providers, perform automated parentage analyses, and manage their DNA data through intuitive interfaces. For organisations seeking a complete solution, the platform can provide an end-to-end system encompassing registry, performance recording, genomic data management, analytical functionality, sales catalogues, and evaluation results publication.

The platform enables breeding organisations to precisely control data access at every level while maintaining workflow flexibility. The backend services power complex analyses through sophisticated algorithms and data management systems, while the frontend provides intuitive interfaces for non-technical users to perform complex genetic analyses. This integrated approach ensures that organisations can maintain complete control over their data and analyses while

benefiting from modern software architecture and tools, facilitating powerful feedback loops between various stakeholders to enhance genetic improvement programs.

MATERIALS AND METHODS

The Helical platform is a species-agnostic full-stack architecture consisting of frontend web interfaces, backend services with databases, and a comprehensive suite of command-line tools. Some modules and tools are catered specifically for diploid species but can be extended as needed for other organisms. The backend module manages databases and analytical procedures through an Application Programming Interface (API), which the frontend and command line tools utilise for data operations and job submissions (Figure 1).

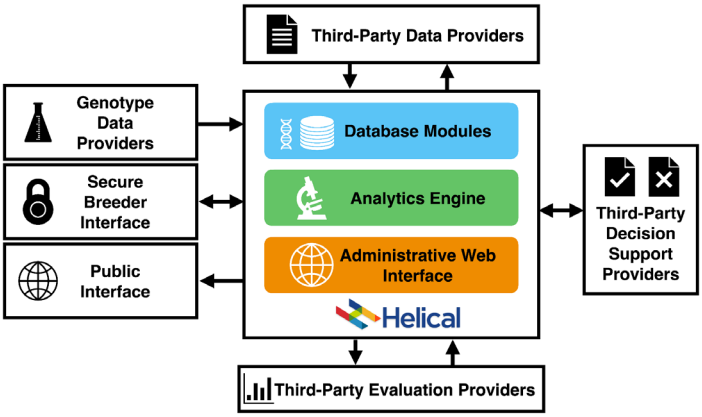

Figure 1. Overview of the Helical platform architecture

The platform is utilised by organisations across many livestock and plant species, with flexible configurations enabling custom terminology to cater for specific industries and use cases.

Frontend interfaces. The platform implements four distinct web portals:

1. Administrative interface for staff to manage system configuration, users and permissions, data templates for import and export, entity management, and analytical workflows including DNA test management and parentage verification or discovery, amongst others.
2. Entity/stakeholder interface for members to manage their animals, view and export DNA and evaluation results, invoices, submit data, and to manage their accounts.
3. Public interface for anonymous users to search animals and estimated breeding values and view published sales catalogues.
4. Technical interface for geneticists, data scientists, or other technical users to access documentation on command line tools and interact with datasets, results, and tools.

Secure backend systems. A web server exposes an API and core database modules including:

- Configuration: business logic and system settings.
- Pedigree and performance: animal data fields.
- DNA: storage/retrieval of sample information and SNP data from various panels.
- Entity management: stakeholder and user data.
- Analysis: various automated and semi-automated processing of pedigree, performance, and genotype data including parent verification and discovery, rigorous genotype sample quality controls, and data export functions.

Optional built-in modules. These extend functionality for:

1) Registry operations (registrations, transfers) including custom business logic and rules.

2) Genetic test results and diagnostics including storing test results for genetic conditions.
3) Performance data for any pre-defined and configured traits.
4) Evaluation results, e.g. estimated breeding values, progeny differences (EBVs/EPDs), accuracies or reliabilities and indexes from one or more evaluation providers.
5) Microsatellite data storage and analysis.

Command line module. The command line tools provide a suite of utilities for technical analyses, data management, and custom pipeline creation. This includes built-in core capabilities for in-house genetic/genomic analyses including single-step evaluations and variance component estimation (heritability estimation), parentage verification and discovery, genotype imputation, DNA data interrogation and analysis, and high-performance linear algebra tools. Helical is routinely used in breeding company scale datasets as well as the largest national and international implementations encompassing tens of millions of animals. The command-line module integrates with the platform's backend API to enable trivial import and export of data stored in Helical database instances. This allows seamless incorporation of any external software tools or pipelines, including enrichment of data from third party systems. Technical users can leverage these capabilities to implement custom analysis pipelines while maintaining version control and reproducibility.

Integrations. The system routinely integrates with third parties:

1) To provide sample IDs and source genotype calls from Neogen, Weatherbys Scientific, Zoetis, and other genotype providers for DNA testing with demonstrated capability to store >1,000,000 genotyped samples on densities up to 700K.
2) To supply pedigree, performance and/or genotype data for national and international evaluations including IGS[1], NZAEL[2], PACE[3], LCC[4], and BREEDPLAN.
3) To display evaluation results and sale catalogues.
4) To produce invoices on Xero software for accrued charges.
5) For mate selection such as provided by MateSel (Kinghorn 2011).

The system is wholly packaged in a single cross-platform executable available under paid license. Because the analytical procedures relating to genomic evaluations exist alongside the backend and frontend modules, there exists a wide scope to surface advanced analyses into the web interfaces, allowing an unlimited amount of extensibility to provide advanced technical functionality to users through secure and accessible web applications.

Any data can be bulk imported or exported in user-customised text formats, e.g. comma separated values. This includes derived data fields generated on the fly via user-defined JavaScript functions. Both the administrative and breeder portal can facilitate such data exchange.

RESULTS AND DISCUSSION

The Australian Wagyu Association's (AWA) implementation of Helical demonstrates the platform's ability to enable sophisticated genomic operations while facilitating valuable data sharing between stakeholders. AWA utilizes Helical as their "One-stop DNA shop," leveraging its ability to integrate with multiple genotyping providers (Neogen and Zoetis) across their international membership base spanning Australia, New Zealand, USA, UK, Canada, and China. The platform manages all aspects of DNA testing from submission through to results delivery, including automated parentage discovery, genetic condition reporting, and trait predictions.

[1]International Genetic Solutions run by American Simmental Association
[2]New Zealand Animal Evaluation Limited, a DairyNZ subsidiary running the NZ dairy evaluation
[3]Pan-American Cattle Evaluation run by the American Hereford Association
[4] Leachman Cattle Company $Profit™ evaluation

AWA's implementation also showcases how Helical's genomic capabilities can drive commercial value through products like the Wagyu Feeder Check (Reverter *et al*. 2023), which ranks animals on five feedlot growth and carcass traits. Commercial producers use these predictions for culling and feeding lot decisions for fullblood and F1 crossbred animals. Helical's automated chip-agnostic parent discovery capabilities create a critical link between commercial animals and AWA registered sires. Producers are credited by AWA for submitting carcass data, which enhances prediction accuracy for the entire membership. This data sharing cooperation enables a powerful data flywheel where commercial producers benefit from valuable decision support tools while the AWA breeding program gains improved carcass trait predictions for registered sires.

AWA utilises the platform's genomic analysis tools to compute measures of inbreeding and diversity, including inbreeding from traditional pedigree coefficients, runs of homozygosity (ROH), or diversity assessed by genomic relationships. These analyses provide members with sophisticated insights to aid breeding decisions, demonstrating how maintaining sovereignty over genomic data allows organisations to implement new tools that create ongoing value from their historical genotype investment. The platform's robust quality control pipelines and flexible analysis framework ensure that genomic data collected over time, even representing different chip densities, manifests, genotyping technologies or service providers, can be leveraged for increasingly sophisticated applications as new methodologies and DNA testing technologies emerge.

The success of the AWA implementation, along with Helical's adoption by numerous organisations across beef and dairy cattle, sheep, dogs, horses, and plants globally, demonstrates how the platform's flexible and agile architecture can be adapted to diverse breeding contexts while maintaining robust genomic capabilities and data sovereignty.

CONCLUSION

The Helical platform is a modern full-stack architecture that provides a future-proofed foundation for breeding program data. It can be deployed as a comprehensive breeding database system or as a standalone genomic database integrating with existing systems. Combined with robust data sharing frameworks, sophisticated genomic analysis capabilities, intuitive interfaces for DNA data management, robust backend services for real-time analytics, and extensive command-line tools, the platform provides organisations complete and agile ownership of their data while maintaining freedom to work with multiple service providers, ensuring they can always access the most competitive pricing and service offerings.

The platform breaks down traditional data silos for the benefit of all stakeholders. Organisations gain not only the freedom to choose and switch between genotyping providers, but also retain permanent access to their historical genomic data and sophisticated analysis tools. This data sovereignty enables them to continue extracting new value from their genomic investment through time, whether through automated parentage discovery, implementation of new trait predictions, or creation of novel data sharing arrangements that benefit both seedstock and commercial operations.

As breeding programs continue to evolve, the ability to independently manage genomic data while maintaining competitive vendor relationships or adopt new technologies will become increasingly critical. This independence ensures organisations can continue to leverage their accumulated genomic data in new ways as technology advances, maximising the long-term return on their breeding program data investment.

REFERENCES

Kinghorn B.P. (2011) *Gen. Sel. Evol.* **43**: 1.
Reverter A., Li Y., Alexandre P.A., Dominik S., *et al*. (2023) *Anim. Prod. Sci.* **64**: AN232446.