# BrAPI v2 - A unified framework for data integration and collaboration for breeding and genetic resources

## Authors

**Peter Selby**[1]*, **Rafael Abbeloos**[2], **Anne-Francoise Adam-Blondon**[3,4], **Francisco J. Agosto-Pérez**[1], **Michael Alaux**[3,4], **Isabelle Alic**[5], **Khaled Al-Shamaa**[6], **Johan Steven Aparicio**[7], **Jan Erik Backlund**[8], **Aldrin Batac**[8,9], **Sebastian Beier**[10,11], **Gabriel Besombes**[5], **Alice Boizet**[12,13], **Matthijs Brouwer**[14], **Terry Casstevens**[15], **Arnaud Charleroy**[5], **Keo Corak**[16], **Chaney Courtney**[17], **Mariano Crimi**[8], **Gouripriya Davuluri**[18], **Kauê de Sousa**[19], **Jeremy Destin**[3,4], **Stijn Dhondt**[2], **Ajay Dhungana**[20], **Bert Droesbeke**[21], **Manuel Feser**[18,22], **Mirella Flores-Gonzalez**[23], **Valentin Guignon**[19], **Corina Habito**[8], **Asis Hallab**[10,24], **Jenna Hershberger**[17], **Puthick Hok**[25], **Amanda M. Hulse-Kemp**[16], **Lynn Carol Johnson**[15], **Sook Jung**[26], **Paul Kersey**[27], **Andrzej Kilian**[25], **Patrick König**[18], **Suman Kumar**[18], **Josh Lamos-Sweeney**[1], **Laszlo Lang**[24], **Matthias Lange**[18], **Marie-Angélique Laporte**[19], **Taein Lee**[26], **Erwan Le Floch**[3,4], **Francisco López**[28], **Brandon Madriz**[29], **Dorrie Main**[26], **Marco Marsella**[28], **Maud Marty**[3,4], **Célia Michotey**[3,4], **Zachary Miller**[15], **Iain Milne**[30], **Lukas A. Mueller**[23], **Moses Nderitu**[31], **Pascal Neveu**[5], **Nick Palladino**[32], **Tim Parsons**[32], **Cyril Pommier**[3,4], **Jean-François Rami**[12,13], **Sebastian Raubach**[30], **Trevor Rife**[17], **Kelly Robbins**[1], **Mathieu Rouard**[33], **Joseph Ruff**[27], **Guilhem Sempéré**[34,35], **Romil Mayank Shah**[36], **Paul Shaw**[30], **Becky Smith**[30], **Nahuel Soldevilla**[8,9], **Anne Tireau**[5], **Clarysabel Tovar**[8,9], **Grzegorz Uszynski**[25], **Vivian Bass Vega**[24], **Stephan Weise**[18], **Shawn C. Yarnes**[32], **The BrAPI Consortium**[37].

[1] Cornell University

[2] VIB Agro-Incubator

[3] Université Paris-Saclay, INRAE, BioinfOmics, Plant Bioinformatics Facility, 78026, Versailles, France

[4] Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France

[5] MISTEA, University of Montpellier, INRAE, Institut Agro, Montpellier, France

[6] ICARDA

[7] Alliance Bioversity-CIAT

[8] Integrated Breeding Platform

[9] Leafnode LLC

[10] Institute of Bio- and Geosciences (IBG-4: Bioinformatics), CEPLAS, Forschungszentrum Jülich GmbH, Wilhelm Johnen Straße, 52428 Jülich, Germany

[11] Bioeconomy Science Center (BioSC), Forschungszentrum Jülich GmbH, 52428 Jülich, Germany

[12] CIRAD, UMR AGAP Institut, Montpellier, France.

[13] AGAP Institut, CIRAD, INRAE, Institut Agro, Université de Montpellier, Montpellier, France.

[14] Wageningen University and Research

[15] Buckler Lab and Institute for Genomic Diversity, Cornell University

[16] USDA-ARS Genomics and Bioinformatics Research Unit

[17] Clemson University

[18] Leibniz Institute of Plant Genetics and Crop Plant Research

[19] Digital Inclusion, Bioversity International, Montpellier, France

[20] Louisiana State University (LSU)

[21] VIB Data Core

[22] Graduate School DILS, Bielefeld Institute for Bioinformatics Infrastructure (BIBI)

[23] The Boyce Thompson Institute

[24] Bingen Technical University of Applied Sciences, Berlinstraße 109, 55411 Bingen am Rhein, Germany

[25] Diversity Arrays Technology (DArT)

[26] Department of Horticulture, Washington State University

[27] Royal Botanic Gardens, Kew

[28] International Treaty on Plant Genetic Resources for Food and Agriculture, FAO

[29] MrBot Software Solutions, Cartago, Costa Rica

[30] The James Hutton Institute

[31] SEQART AFRICA

[32] Breeding Insight, Cornell University

[33] Bioversity International, Parc Scientifique Agropolis II, 34397 Montpellier, France

[34] CIRAD, UMR INTERTRYP, Montpellier, France INTERTRYP, Univ Montpellier, CIRAD, IRD, Montpellier, France French Institute of Bioinformatics (IFB)

[35] South Green Bioinformatics Platform, Bioversity, CIRAD, INRAE, IRD, Montpellier, France

[36] North Carolina State University

[37] The BrAPI Consortium

* Corresponding author E-mail: ps664@cornell.edu (PS)

## Abstract

Recent advancements in phenotyping techniques, genotyping technologies, and prediction approaches are facilitating increased genetic gains in breeding programs, helping researchers and breeders develop better crops at a faster rate. However, these advancements have also led to an overwhelming torrent of disconnected data. The successful implementation of these methods depends on proper data management, which is particularly challenging due to the need to integrate datasets across various types, formats, and sources. To address these challenges, the Breeding API (BrAPI) project was created and has been adopted as an accepted standard data model in the breeding domain. The BrAPI project is an international effort that is enabling more efficient data management through the development of interoperable databases and tools that can be used to share and interpret breeding-related data. This community driven standard is software agnostic, open-source, and can be used by anyone interested in plant breeding, phenotyping, germplasm, genotyping, and agronomy data management. This manuscript presents an overview of the BrAPI project, the substantial growth of the data standard, and the wide variety of BrAPI-compatible, community-built tools for breeding and research. This manuscript provides an overview of the BrAPI project, highlighting the significant progress made in the development of the data standard and the expansion of its community. It also presents a showcase of the wide variety of BrAPI-compatible tools that have been built to enhance breeding and research activities, demonstrating how the project is advancing agricultural innovation and data management practices.

## Author Summary

Population growth and the impacts of climate change are placing increasing pressure on global agriculture and breeding programs. To meet the growing demand for food and resources while coping with the unpredictability of changing climates, extraordinary efforts are required to boost the efficiency of these systems. This increase in efficiency will in part come from better computer technology and better data management. The ability to share data and software is crucial to get better technology with fewer resources, allowing scientists to build on each others work more easily. Creating an international data standard, such as BrAPI, within the breeding and agricultural domain allows for easy data sharing and software interoperability. The BrAPI community combines the expertise of breeders, computational biologists, software engineers, and data scientists to build this agreed upon data standard. The community actively encourages collaboration between research groups, creating a network of shared technology and knowledge. This spirit of collaboration not only accelerates progress but also helps to use resources more efficiently, reducing duplication of effort and enabling research groups to focus on new innovations.

# Introduction

Breeding programs aim to deliver improved lines or cultivars, the most fundamental input for farming, and are thus foundational for maintaining a productive agricultural system amidst the pressing challenges of climate change. Breeding efforts are time- and resource-intensive, with progress dependent on efficient program logistics and accurate selection decisions. While breeding programs can benefit from modern and emerging breeding techniques like genomic selection, machine learning, and high throughput phenotyping, the successful implementation of these methods depends on the ability to efficiently collect, manage, and analyze large volumes of carefully curated genomic and phenomic data [1]. Extracting actionable knowledge from these complex datasets is time-consuming, often prohibiting the adoption of new methods, especially by under-resourced breeding programs. To facilitate the collection, management, and analysis of these datasets, it is essential to transition to digital tools. Historically, independent applications were designed to address specific problems, but in many cases, this led to separate software solutions for each breeding program task and created data silos.

The [Breeding Application Programming Interface (BrAPI)](#) is a standardized, representational state transfer (REST), web service, application programming interface (API) specification for breeding and related agricultural data [2]. Since the project inception in 2014, BrAPI has become an essential part of the digital infrastructure for plant breeding, providing a domain-specific open data standard tailored to the needs of plant breeding and genetics projects. BrAPI enables interoperability between breeding software platforms, allowing groups to seamlessly share data and software tools both within and across breeding programs. It eases the merging of datasets of different types and provides access to shared trait ontologies, phenotypic data, genotypes, seed inventories, and other essential components for collaborative breeding efforts.

Since its first publication in 2019 [2], BrAPI has seen a significant increase in community services, compatible tools, and participating organizations. The community has organized numerous hackathons to evolve the specification, resulting in continuous improvements and enhancements. This report includes a short technical description of the standard and a showcase of the applications, services, and tools available from the BrAPI community. It is the intention of this manuscript to demonstrate the value of BrAPI to the wider scientific community as an effective and efficient means to collaborate and exchange data.

## How it works

An API is a technical connection between two pieces of software. Just as a graphical user interface (GUI) or a command line interface (CLI) allows a human user to interact with a piece of software, an API allows one software application to interact with another. A REST-style (or RESTful) web service is a type of API commonly used in modern web infrastructure. REST is a technical architecture that describes the stateless transmission of data between applications. Typically, RESTful web service APIs are implemented using the standard HTTP protocol that most of the modern internet is built upon. These implementations generally use JavaScript Object Notation (JSON) to represent the data being transferred. Both HTTP and JSON are programming language agnostic, very stable, and highly flexible. This means BrAPI can be implemented in almost any piece of software and can solve a wide range of use cases.

Data repositories and service providers that are BrAPI compatible have mapped their internal data structures to the BrAPI standard models, allowing them to share data with the outside world in a standardized format. Similarly, they can accept new data from external sources and automatically map the new data to their existing database. Client application developers can take advantage of this

standardization by building tools and connectors that integrate with all BrAPI-compatible data repositories. Visualization, reporting, analytics, data collection, and quality control tools can be built once and shared with other organizations that follow the standard. This type of BrAPI-compatible, easily sharable tool is often referred to as a BrAPP, meaning BrAPI Application. BrAPPs are simple tools that are entirely reliant on BrAPI for their data requirements, and often fit on a single web page. A single BrAPP can be easily shared and used by many organizations and systems, as long as those organizations have the required BrAPI endpoints available. As the number of BrAPI-compatible databases, tools, and organizations grows, so does the value of implementing the standard into any given application.

## Project Updates

Over its lifetime, the BrAPI project has grown and changed substantially. The total size of the specification has almost quadrupled since the release of version v1.0 in 2017, increasing from 51 endpoints in v1.0 to 201 endpoints in v2.1. Because of this growth, the specification documents were reorganized into four modules: BrAPI-Core, BrAPI-Phenotyping, BrAPI-Genotyping, and BrAPI-Germplasm. Figure 1 is a simplified domain map of the whole BrAPI data model, showing what kinds of data are defined in each module. While early versions of the specification focused on read-only phenotype data, the specification now has representation from most of the major concepts related to breeding. The newest specification has also been updated to be internally consistent, easier to navigate, and allow for read, write, and update capabilities.
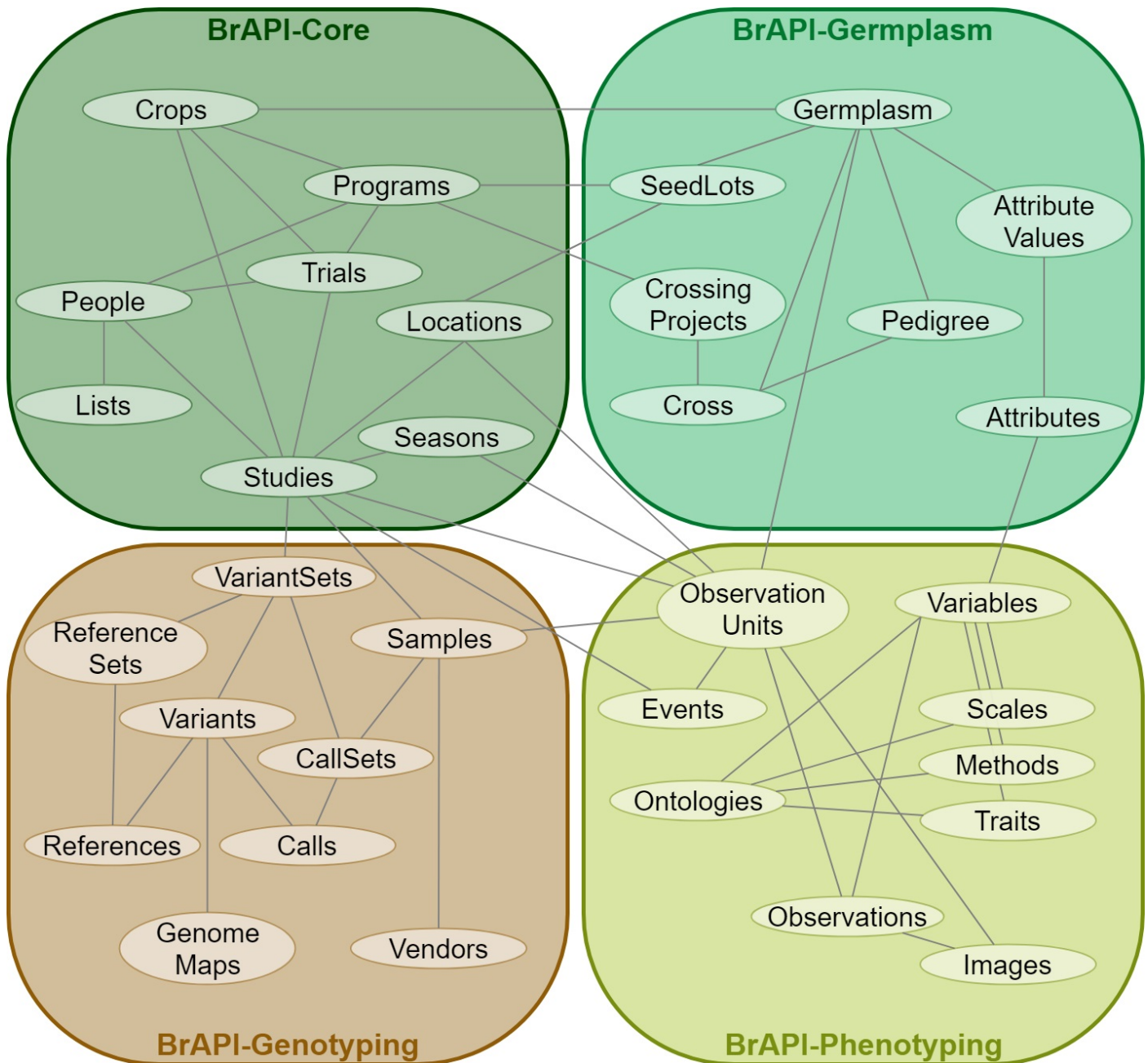
**Figure 1:** A simplified domain map of the whole BrAPI data model, divided into organizational modules. A more detailed Entity Relationship Diagram (ERD) is available on brapi.org.

As BrAPI has matured, so have the tools, services, and libraries that work with the specification. Each new version is released with a change log to guide developers as they upgrade, an Entity Relationship Diagram (ERD) to visually describe the data model, and a JSON Schema data model to be used for automated development efforts. For groups using Java, Java Script, Python, R, or Drupal, community-maintained libraries are available with full BrAPI implementations ready to be integrated into existing code. The BrAPI Test Server is updated to support every version of the specification for testing purposes. Finally, there are resource pages on the [project website](#) that showcase BrAPI-compatible applications and data resources available in the community.

## Community Growth

The international BrAPI Community consists of software developers, biologists, and other scientists working on BrAPI related projects and data sources. This community sustains the BrAPI project, builds implementations, maintains development tools, and provides input to enhance the specification. As the project has grown, so too has the community. The BrAPI project started in June 2014 with less than ten people coming together to discuss the idea and has since grown to more than 200 members.

The BrAPI Hackathons are a major staple of the BrAPI community [3]. Twice a year, the community gathers in person or virtually to discuss the specification and collaborate on BrAPI-related projects. These events have proven to be vital to the long-term growth of the community; for some organizations, the hackathon is the only time during the year when they can collaboratively work on BrAPI projects.

## Results

Below are a number of short success stories from the BrAPI community. These tools, applications, and infrastructure projects serve as another indicator of community growth and success over the past 5 years. These stories clearly illustrate all the different ways the BrAPI standard can be used productively and in practice. Figure 2 contains a summary of many of the currently available BrAPI-compliant tools, and each will be further described below.

| Tool | Core | | | | Phenotyping | | | | Genotyping | | | | Germplasm | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Generation/Collection | Storage | Visualization | Analysis | Generation/Collection | Storage | Visualization | Analysis | Generation/Collection | Storage | Visualization | Analysis | Generation/Collection | Storage | Visualization | Analysis |
| Field Book | | | | | ✓ | ✓ | | | | | | | | | | |
| GridScore | | | | | ✓ | ✓ | ✓ | | | | | | | | | |
| ClimMob | ✓ | ✓ | | | ✓ | ✓ | | | | | | | | | | |
| ImageBreed | | | | | | ✓ | ✓ | ✓ | | | | | | | | |
| PHIS | | ✓ | | | ✓ | ✓ | | ✓ | | | | | | ✓ | | |
| PIPPA | | ✓ | | | ✓ | ✓ | | | | | | | | ✓ | | |
| Trait Selector BrAPP | | | | | | | ✓ | | | | | | | | | |
| DArT Sample Submission | | | | | | | | | ✓ | | | | | | | |
| DArTView | | | | | | | | | | | ✓ | ✓ | | | | |
| DivBrowse | | | | | | | | | | ✓ | ✓ | ✓ | | | | |
| Flapjack | | | | | | | ✓ | | | | ✓ | ✓ | | | | |
| GIGWA | | | | | | | | | | ✓ | ✓ | ✓ | | | | |
| PHG | | | | | | | | | ✓ | ✓ | | ✓ | | | | |
| AGENT Information System | | ✓ | | | ✓ | | | | | ✓ | ✓ | ✓ | | ✓ | | |
| FLORILÈGE | | | | | | | | | | | | | | ✓ | ✓ | |
| GLIS | | | | | | | | | | | | | ✓ | ✓ | | |
| Helium | | | | | ✓ | ✓ | | | | | | | | ✓ | ✓ | |
| MGIS | ✓ | ✓ | | | ✓ | ✓ | | | | | | | ✓ | ✓ | | ✓ |
| BIMS | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | | | | | ✓ | ✓ | |
| BMS | ✓ | ✓ | ✓ | | ✓ | ✓ | | ✓ | | ✓ | | | | ✓ | ✓ | |
| Breedbase | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | | | ✓ | ✓ | ✓ |
| DeltaBreed | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | ✓ | |
| FAIDARE | | | ✓ | | | ✓ | | | | | ✓ | | | | ✓ | |
| Germinate | ✓ | ✓ | ✓ | | ✓ | ✓ | | | | ✓ | | | | ✓ | ✓ | |
| G-Crunch | | | | ✓ | | | | ✓ | | | | ✓ | | | | ✓ |
| QBMS | | | | ✓ | | | | ✓ | | | | ✓ | | | | ✓ |
| Mr. Bean | | | | | | | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | ✓ |
| SCT | | | | | | | | | | | | | | | ✓ | ✓ |
| ShinyBrAPPs | | | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | ✓ | | | ✓ | ✓ |

**Figure 2:** A summary of all the tools described below and the general areas each tool is designed to handle. The "Generation/Collection" column indicates that an application is used to input or create new data. The "Storage" column indicates the tool stores that type of data. The "Visualization" column indicates that application has a way of presenting data to a human user. The "Analysis" column indicates the tool does some calculation to provide new insight.

# Phenotyping

Phenotyping is fundamental to plant breeding and genetics research, providing the accurate, high-quality data needed for downstream analyses and selection decisions. Effective phenotyping requires

a thorough understanding of both biological research questions and operational data gathering techniques to ensure successful outcomes. Collected data, subsequent analyses, and data visualizations all impact and shape downstream experiments. The BrAPI specification supports phenotypic data throughout the entire breeding pipeline, including collection, analyses, publication, and archiving. The BrAPI community has developed several BrAPI-compatible tools to facilitate data standardization, efficient storage, and curation of phenotypic data and trait metadata. Ongoing development efforts are creating tools to manage images and other high-throughput phenotypic data sources, further enhancing the precision and efficiency of plant breeding research. By supporting the collection and storage of phenotypic data accurately and efficiently, BrAPI compatible tools (as outlined below) simplify the conversion of phenotypes into actionable insights that are necessary to help digitize and boost modern breeding and genetics research programs. The following set of BrAPI-compatible tools were developed to support some aspect of the phenotyping process.

## Field Book

Data from plant breeding and genetics experiments has traditionally been collected using pen and paper, but this approach often results in transcription errors and delayed analysis. Field Book [4], a highly-customizable Android app, was developed to help scientists digitize and organize their phenotypic data as measurements and images are collected. This effectively improves data collection speed, reduces errors, and enables larger and more robust breeding populations and data sets.

Field Book has added support for BrAPI to streamline data transfer to and from BrAPI-compatible servers. This improvement has removed the need to manually transfer data files, simplifies data exchange between these systems, and reduces the opportunities for human error and data loss.

## GridScore

GridScore [5] is a web-based application for recording phenotypic observations that harnesses mobile devices to enrich the data collection process. The GridScore interface closely mirrors the look and feel of printed field plans, creating an intuitive user experience. GridScore performs a wide range of functions, including data validation, data visualization, georeferencing, multi-platform support, and data synchronization across multiple devices. The application's data collection approach employs a top-down view onto the trial and offers field navigation mechanisms using barcodes, QR codes, or guided walks that take the data collector through the field in one of 16 pre-defined orders.

BrAPI has further increased the value of GridScore by integrating it into the overarching plant breeding workflow. Trial designs and trait definitions can be imported into GridScore using BrAPI and a finalized trial can be exported via BrAPI to any compatible database.

## ClimMob

ClimMob [6] is a software suite for an alternative research paradigm in experimental agriculture. In traditional breeding, a small number of researchers design complicated trials in search of the best solutions for few target environments. ClimMob applies the principles of citizen science and choice experiments to scale the data collection process, mostly in the form of accession rankings. Although these data may not be as detailed as from a centralized experiment, they can inform decisions across a wide range of environments with increased external validity. Applications of ClimMob include crop variety testing, evaluating agronomic practices, and investigating climate resilience strategies. The platform supports experiment design, data collection through mobile apps, and data analysis to provide actionable insights.

ClimMob uses BrAPI to retrieve curated germplasm information from breeding databases for trial design, subsequently enabling the automatic upload of collected ClimMob-collected data to a central breeding database for long-term storage and analysis. Analyzed data can also be pushed from ClimMob to breeding databases, providing breeders with insights into the potential adoption of the tested crop varieties.

## ImageBreed

Unoccupied aerial and ground vehicles (UAVs and UGVs) enable the high throughput collection of images and other sensor data in the field, but the rapid processing and management of these datasets are often a bottleneck for breeding programs seeking to deploy these technologies for time-sensitive decision making. ImageBreed [7] is an open-source, BrAPI-compliant image processing tool that supports the routine use of UAVs and UGVs in breeding programs through standardized pipelines. It creates orthophotomosaics, applies filters, assigns plot polygons, and extracts ontology-based phenotypes from raw UAV-collected images. The BrAPI standard is used to push these phenotypes back to a central BrAPI-compliant breeding database where they can be analyzed with other experiment data. The ImageBreed team has collaborated with others in the community to enhance the BrAPI image data standards, which it uses to upload raw images to a central breeding database, or any other BrAPI-compatible long term storage service.

## PHIS

PHIS [8], the Hybrid Phenotyping Information System, is an ontology-driven information system based on semantic web technologies and the OpenSILEX framework. PHIS is deployed in several field and greenhouse platforms of the French national PHENOME and European EMPHASIS infrastructures. It manages and collects data from basic phenotyping and high-throughput phenotyping experiments on a daily basis. PHIS unambiguously identifies the objects and traits in an experiment and establishes their types and relationships via ontologies and semantics.

Since its inception, PHIS has been designed to be BrAPI compliant, encompassing the Core, Phenotyping, and Germplasm BrAPI modules. This enables integration with other BrAPI-compliant systems and platforms, simplifying the exchange of accession and phenotyping data across systems. PHIS is actively integrated with the OLGA genebank accessions management system , and is indexed by the FAIDARE data portal [9]. BrAPI-enabled interoperability promotes a more coherent and efficient approach to the management and use of phenotyping data on various platforms and research initiatives within the European scientific community. BrAPI compliance also ensures that PHIS is compatible with other standards such as the Minimal Information About a Plant Phenotyping Experiment (MIAPPE [10]). By integrating BrAPI requirements into its structure, PHIS strengthens its capacity for interoperability and effective collaboration in the wider context of plant breeding and related fields.

## PIPPA

PIPPA [11] is a data management system used for collecting data from the Weighing, Imaging and Watering Machines (WIWAM) [12], which are a range of automated high-throughput phenotyping platforms. These platforms have been deployed by research institutes and commercial breeders across Europe. They can be set up in a variety of configurations with different types of equipment including weighing scales, cameras, and environment sensors. The software features a web interface with functionality for setting up new experiments, planning imaging and irrigation treatments, linking metadata (genotype, growth media, manual treatments) to pots, and importing, exporting, and visualizing data. It also supports the integration of image analysis scripts and connects to a compute cluster for job submission.

To share the phenotypic data from PIPPA experiments linked to publications, an implementation of BrAPI v1.3 was developed which allowed read only access to the data in the BrAPI standardized format. This server was registered on FAIDARE [9], allowing the data to be found alongside data from other BrAPI-compatible repositories.

Throughout its development, the PIPPA project has adhered to guidelines set forth by BrAPI and the MIAPPE [10] scientific standard. Current efforts are focused on delivering a public BrAPI v2.1 endpoint and increasing the availability of public high-throughput datasets via BrAPI.

### Trait Selector BrAPP

The Trait Selector BrAPP is a JavaScript-based application used to visually search and select traits from an ontology. The Trait Selector employs a visual aid, an image of a plant, to connect plant anatomy with relevant trait ontology terms. Instead of scrolling through a long list of possible traits, the user can click on pieces of the image to show the traits associated with specific plant structures. The Trait Selector BrAPP can be used to quickly find specific traits or to identify accessions that have a specific phenotype of interest.

The Trait Selector BrAPP has been successfully added to Cassavabase[13] and MGIS[14], and it can be integrated into any website or system with a BrAPI-compatible data source. A breeding database would need to only implement the BrAPI endpoints for Traits, Observations, and Variables, while a genebank would require only Traits and Germplasm Attributes.

## Genotyping

Genotyping has become a cornerstone of most breeding processes, but managing the data can be challenging. Understanding different genotyping protocols for various crops is crucial due to the unique genetic structures of each species. Techniques such as single nucleotide polymorphism (SNP) genotyping, Genotyping-by-Sequencing (GBS), SSRs, Whole Genome Sequencing (WGS), and array-based genotyping each offer specific advantages depending on the crop and research objectives. BrAPI supports genotypic data by utilizing existing standards such as the variant call format (VCF) [15] and the GA4GH Variants schema [16]. The BrAPI community has developed compatible tools for storing, searching, visualizing, and analyzing genotypic data, making it easier to integrate and utilize this information in breeding programs. Mastery of the various genotyping protocols ensures efficient and effective breeding, while BrAPI-compliant tools streamline data management and analysis, enhancing the ability to make data-driven decisions in developing superior crop varieties.

### DArT Sample Submission

The Diversity Arrays Technology (DArT) genotyping lab is heavily used worldwide when it comes to plant genotyping. With over 1200 available organisms and species, a client base on every continent, and many millions of samples processed, DArT provides services for several generic and bespoke genotyping technologies and solutions. Processes of sample tracking and fast data delivery are at the core of the ordering system developed at DArT. The ordering system is tightly integrated with DArTdb (DArT's custom LIMS operational system), which drives laboratory, quality, and analytical processes.

Diversity Arrays Technology has been a part of the BrAPI community since its inception. DArT developers have worked with the BrAPI community contributing to various aspects of the API specification. One key aspect was establishing a standard API for sending sample metadata to the lab for genotyping. This solution eliminates much of the human error involved with sending samples to an external lab, and also allows for an automated process of sample batch transfers. The current

implementation also allows for an order status verification, automated data discovery, and data downloads. Data are delivered as standard data packages with self-describing metadata.

The current BrAPI implementation at DArT is in production and it is compatible with the BrAPI v2.1 specification. Further details about DArT's ordering system can be found at [DArT Ordering System](#) and also at [DArT Help](#).

## DArTView

[DArTView](#) is a desktop application for marker data curation via metadata filtering. DArTView enables genotype variant data visualization designed such that users can easily identify trends or correlations within their data. The primary goal of the tool is to overcome tedious manual calculation of marker data through common spreadsheet applications like Excel. Users are able to import marker data from csv files, but DArTView has been recently enhanced to be BrAPI compatible. BrAPI provides a consistent data standard across databases and data resources, which allows DArTView to use any BrAPI-compatible server as an input data source. DArTView's compatibility with BrAPI also ensures easy integration with other tools and pipelines that would use DArTView for marker filtering and exploration.

Initially developed by DArT, the tool is gaining popularity within the breeding community, especially in Africa. Future releases will focus on enhancing the BrAPI compatibility, making it accessible to more breeders and researchers. A web enabled version of DArTView is in development. This new version will allow for further collaboration opportunities with other interested partners who would like to integrate it as part of their pipelines.

## DivBrowse

[DivBrowse](#) [17] is a web platform for exploratory data analysis of large genotyping studies. The software can be run standalone or integrated as a plugin into existing web portals. At its core, DivBrowse combines the convenience of a genome browser with features tailored to germplasm diversity analysis. DivBrowse provides visual access to VCF files obtained through genotyping experiments and can handle hundreds of millions of variants across thousands of samples. It is able to display genomic features such as nucleotide sequence, associated gene models, and short genomic variants. DivBrowse also calculates and displays variant statistics such as minor allele frequencies, the proportion of heterozygous calls, and the proportion of missing variant calls. Dynamic principal component analyses can be performed on a user-specified genomic area to provide information on local genomic diversity. DivBrowse also has an interface to BLAST+ tools [18] installed on Galaxy servers [19], which can be used to directly access genes or other genomic features from results of custom BLAST query. DivBrowse employs the BrAPI-Genotyping module to serve genotypic data as a BrAPI endpoint and to get genotypic data from other BrAPI endpoints.

## Flapjack

[Flapjack](#) [20] is a multi-platform desktop application for data visualization and breeding analysis (e.g., pedigree verification, marker-assisted backcrossing and forward breeding) using high-throughput genotype data. Data can be imported into Flapjack from any BrAPI-compatible data source with genotype data available. [Flapjack Bytes](#) is a smaller, lightweight, and fully web-based counterpart to Flapjack that can be easily embedded into a database website to provide similar visualizations online. Traditionally supporting its own text-based data formats, Flapjack's use of BrAPI has streamlined the end-user experience for data import. Work is underway to determine the best methods to exchange analysis results using future versions of the API.

## Gigwa

[Gigwa](#) is a Java EE web application providing a means to centralize, share, finely filter, and visualize high-throughput genotyping data [21]. Built on top of MongoDB, it is scalable and can support working smoothly with datasets containing billions of genotypes. It is installable as a Docker image or as an all-in-one bundle archive. It is straightforward to deploy on servers or local computers and has thus been adopted by numerous research institutes from around the world. Notably, Gigwa serves as a collaborative management tool and a portal for exploring public data for genebanks and breeding programs at some CGIAR centers [22]. The total amount of data hosted and made widely accessible using this system has continued to grow over the last few years.

The Gigwa development team has been involved in the BrAPI community since 2016 and took part in designing the genotype-related section of the BrAPI standard. Gigwa's first BrAPI-compliant features were designed for compatibility with the Flapjack visualization tool [20]. Over time, Gigwa has established itself as the first and most reliable implementation of the BrAPI-Genotyping module. Local collaborators and external partners used it as a reference solution to design a number of tools taking advantage of the BrAPI-Genotyping features (e.g., [BeegMac](#), [SnpClust](#), [QBMS](#)).

Some use-cases require Gigwa to also consume data from other BrAPI servers. This requirement led to the implementation of BrAPI client features within Gigwa. A close collaboration was established with the Integrated Breeding Platform team and their widely used Breeding Management System (BMS). This collaboration means both applications are now frequently deployed together; Gigwa pulling germplasm or sample metadata from BMS, and BMS displaying Gigwa-hosted genotypes within its own UI.

## PHG

The [Practical Haplotype Graph (PHG)](#) is a graph-based computational framework that represents large-scale genetic variation and is optimized for plant breeding and genetics [23]. Using a pangenome approach, each PHG stores haplotypes (the sequence of part of an individual chromosome) to represent the collective genes of a species. This allows for a simplified approach for dealing with large scale variation in plant genomes. The PHG pipeline provides support for a range of genomic analyses and allows for the use of graph data to impute complete genomes from low density sequence or variant data.

Users can access the haplotype data either with direct calls to the PHG embedded server or indirectly using the rPHG library from an R environment. The PHG server accepts BrAPI queries to return information on sample lists and the variants used to define the graph's haplotypes. In addition, PHG users utilize the BrAPI variant sets endpoint query to return links to VCF files containing haplotype data. Work on the PHG is ongoing and it is expected to support additional BrAPI endpoints that allow for fine tuned slicing genotypic data in the near future.

# Germplasm Management

Germplasm data management is crucial due to the vast quantity of new accessions, variants, and lines created yearly. Germplasm is the basis of variation on which breeders rely to upgrade and optimize their breeding programs. This is essential at any scale including individual breeding programs, national initiatives, and international collaborations. BrAPI supports the transmission of germplasm passport data, pedigree trees, and crossing metadata. The BrAPI community has developed compliant tools for storing, searching, and visualizing this metadata, facilitating efficient management. Additionally, there are plans to establish federated networks of genebank data connected via BrAPI, enhancing global accessibility and collaboration in germplasm management.

## AGENT

The aim of the AGENT project, funded by the European Commission, is to develop a concept for the digital exploitation and activation of plant genetic resources (PGRs) throughout Europe [24]. In the global system for ex situ conservation of PGRs, material is being conserved in about 1750 collections totalling ~5.8 million accessions [25]. Unique and permanent identifiers in the form of DOIs are available for more than 1.7 million accessions via the Global Information System [26] of the International Treaty on Plant Genetic Resources for Food and Agriculture. Each DOI is linked to some basic descriptive data that facilitates the use of these resources, mainly passport data. However, a data space beyond the most basic information is needed that includes genotypic and phenotypic data. This space will help to answer questions about the global biological diversity of plant species, the detection of duplicates, the tracking of provenance for the identification of genetic integrity, the selection of the most suitable material for different purposes, and to support further applications in the field of data mining or AI. In this context, the AGENT project will activate and utilize the PGRs from European ex situ genebanks according to the FAIR principles, and test the resources in practice using two important crops, barley and wheat [27]. Thirteen European genebanks and five bioinformatics centers are working together and have agreed on standards and protocols for data flow and data formats for the collection, integration, and archiving of genotypic and phenotypic data[28].

The BrAPI specification is one of the agreed standards that are detailed in the AGENT guidelines for dataflow [29]. The implemented BrAPI interface enables the analysis of current and historic genotypic and phenotypic information. This will drive the discovery of genes, traits, and knowledge for future missions, complement existing information for wheat and barley, and use the new data standards and infrastructure to promote better access and use of PGR for other crops in European genebanks. The AGENT database backend aggregates curated passport data, phenotypic data, and genotypic data on wheat and barley accessions of 18 project partners. This data is accessible via BrAPI endpoints and explorable in a web portal. Genotyping data uses the DivBrowse [17] storage engine and its BrAPI interface. Soon, the BrAPI implementation will be expanded to enable the integration of analysis pipelines in the AGENT portal, such as the FIGS+ pipeline developed by ICARDA [30]. There is also a plan to integrate the data collected by the AGENT project into the European Search Catalogue for Plant Genetic Resources (EURISCO) [31].

## Florilège

Florilège is a web portal designed primarily for the general public to access public plant genetic resources held by the Biological Resource Centers across France, as part of France's National Research Institute for Agriculture, Food and Environment (INRAE). Through this portal, users can browse accessions from over 50 plant genera, spread across 19 genebanks. It allows users to view available seeds and plant material, including options for ordering material. Florilège provides a centralized access to the various French collections of plant genetic resources available to the public.

Florilège retrieves accession information from several BrAPI-compliant systems. Key among these are OLGA, a genebank accessions management system, and GnpIS, an INRAE data repository for plant genetic resources, phenomics, and genetics [32,33]. Using BrAPI to gather data from these systems reduced development efforts and enabled standardized data retrieval. As a result, BrAPI has become the de facto standard within the French plant genetic resources community for exchanging information. During development, the Florilège team also proposed several enhancements to the BrAPI specifications themselves, such as additional support for Collection objects or improved reference linking, to better accommodate their specific use case.

## GLIS

The [Global Information System (GLIS)](#) on Plant Genetic Resources for Food and Agriculture (PGRFA) of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) is a web-based, BrAPI-compliant global entry point for PGRFA data [26]. It allows users and third-party systems to access information and knowledge on scientific, technical, and environmental matters to strengthen PGRFA conservation, management, and utilization activities. The system and its portal also enable recipients of PGRFA to make available all non-confidential information on germplasm according to the provisions of the Treaty and facilitates access to the results of their research and development.

Thanks to the adoption of Digital Object Identifiers (DOIs) for Multi-Crop Passport Descriptors (MCPD) of PGRFA accessions, the GLIS Portal provides access to 1.7 million PGRFA in collections conserved worldwide. Of these, over 1.5 million are accessible for research, training and plant breeding in the food and agriculture domain.

The Scientific Advisory Committee of the ITPGRFA have repeatedly welcomed efforts on interoperability among germplasm information systems. In this context, the GLIS Portal adopted the BrAPI v1.3 in 2022. Integrating BrAPI among the GLIS content negotiators facilitates queries and the exchange of content for data management in plant breeding. The Portal also offers other protocols (XML, DarwinCore, JSON and JSON-LD) to increase data and metadata connectivity. In the near future, depending on the availability of resources, upgrading to BrAPI v2 is planned.

## Helium

[Helium](#) [34] is a plant pedigree visualization platform designed to account for the specific problems that are unique to plant pedigrees. A pedigree is a representation of how genetically discrete individuals are related to one another and is therefore a representation of the genetic relationship between individual plant lines, their parents and progeny. Plant pedigrees are often used to check for potential genotyping or phenotyping errors, since these errors, by the very nature of Mendelian inheritance, are constrained by the pedigree structure in which they exist [35]. The accurate representation of plant pedigrees, and the ability to pull pedigree data from different data sources is important in plant breeding and genetics. Therefore, ways to visualize and interact this complex data in meaningful ways is critical.

From its [original desktop interface](#), Helium has developed into a web-based visualization platform implementing BrAPI calls to allow users to import data from other BrAPI-compliant databases. The ability to pull data from BrAPI-compliant data sources has significantly expanded Helium's capability and utility within the community. Helium is used in projects ranging in size from tens to tens of thousands of lines and across a wide variety of crops and species. While originally designed for plant data [36] it has also found utility in other non-plant projects [37] highlighting its broad utility. BrAPI also allows Helium to provide direct dataset links to collaborators, allowing the original data to be held with the data provider and utilizing Helium for its visualization functionality. Our current Helium deployment includes example BrAPI calls to a barley dataset at the James Hutton Institute to allow users to test the system and features it offers.

## MGIS

The [Musa Germplasm Information System (MGIS)](#) serves as a comprehensive community portal dedicated to banana diversity, a crop critical to global food security [14]. MGIS offers detailed information on banana germplasm, focusing on the collections held by the CGIAR International Banana Genebank (ITC) [38]. It is built on the Drupal/Tripal technology, like BIMS [39] and Florilège.

Since its inception, MGIS developers have actively participated in the BrAPI community. The MGIS team pushed for the integration of the Multi-Crop Passport Data (MCPD) standard into the

Germplasm module of the API. MCPD support was added in BrAPI v1.3, and MGIS now provides passport data information on ITC banana genebank accessions (with GLIS DOI), synchronized with Genesys. MGIS also enriches the passport data by incorporating additional information from other germplasm collections worldwide. All the germplasm data is available through the BrAPI-Germplasm module implementation. For genotyping data, MGIS integrates with Gigwa [21], which provides a tailored implementation of the BrAPI genotyping module. Furthermore, MGIS supports a set of BrAPI-Phenotyping module endpoints, facilitating the exposure of morphological descriptors and trait information supported by ontologies like the Crop Ontology [40]. MGIS has integrated the Trait Selector BrAPP, and there are use cases implemented to interlink genebank and breeding data between MGIS and the breeding database MusaBase.

## Breeding and Genetics Data Management

While specialty data management is important for some use cases, often breeders want a central repository or access point of critical data. General breeding and genetics data management systems and web portals support some level of phenotypic, genotypic, and germplasm data, as well as trial, equipment, and people management. By enabling BrAPI support, these larger systems can connect with smaller tools and specialty systems to provide more functionality under the same user interface. There are several breeding data management systems developed in the BrAPI community, each with their own strengths.

### BIMS

The Breeding Information Management System (BIMS) [39] is a free, secure, online breeding management system which allows breeders to store, manage, archive, and analyze their private breeding program data. BIMS enables individual breeders to have complete control of their own breeding data along with access to tools such as data import, export, analysis, and archiving for their germplasm, phenotype, genotype, and image data. BIMS is currently implemented in five community databases, the Genome Database for Rosaceae [41], CottonGEN [42], the Citrus Genome Database, the Pulse Crop Database, and the Genome Database for Vaccinium, where it enables individual breeders to import publicly available data. BIMS is also implemented in the public database breedwithbims.org that any breeder can use.

BIMS primarily utilizes BrAPI to connect with Field Book [4], enabling seamless data transfer between data collection and subsequent management in BIMS. Data transfer through BrAPI between BIMS and other resources such as Breedbase [13], GIGWA [21], and the Breeder Genomics Hub [43/] is under development.

### BMS

The Breeding Management System (BMS), developed by the Integrated Breeding Platform (IBP), is a suite of tools designed to enhance the efficiency and effectiveness of plant breeding. BMS covers all stages of the breeding process, with the emphasis on germplasm management and ontology-harmonized phenotyping (i.e. with the Crop Ontology). It also features analytics and decision-support tools. With its focus on interoperability, BMS integrates smoothly with BrAPI, facilitating easy connections with a broad array of complementary tools and databases. Notably, the BMS is often deployed together with Gigwa to fulfill the genotyping data management needs of BMS users.

The brapi-sync tool, a significant component of BMS's BrAPI capabilities, was developed by the IBP and released as a BrAPP for community use. Brapi-sync is designed to enhance collaboration among partner institutes within a network such as Innovation and Plant Breeding in West Africa (IAVAO). The tool enables the sharing of germplasm and trial metadata across BrAPI-enabled systems. It helps

overcome traditional barriers to collaboration, ensuring data that was once isolated within specific programs or platforms can now be easily shared, integrated, and synchronized.

Additionally, brapi-sync improves data management by maintaining links to the original source of each entity it transmits. This retains the original context of the data and establishes a traceability mechanism for accurate data source attribution and verification. Such practices are crucial for maintaining data integrity and fostering trust among collaborative partners, ensuring access to accurate, reliable, and current information.

## Breedbase

Breedbase is a comprehensive, open-source, breeding data management system [13,44] that implements a digital ecosystem for all breeding data, including trial data, phenotypic data, and genotypic data. Data acquisition is supported through data collection apps such as Fieldbook [4], Coordinate, and InterCross, as well as through drone imagery, Near Infra-Red Spectroscopy (NIRS), and other technologies. Search functions, such as the Search Wizard interface, provide powerful query capabilities. Various breeding-centric analysis tools are available, including mixed models, heritability, stability, principal component analysis (PCA), and various clustering algorithms. The original impetus for creating Breedbase was the advent of new breeding paradigms based on genomic information such as genomic prediction algorithms [45] and the accompanying data management challenges. Thus, complete genomic prediction workflow is integrated in the system.

The BrAPI interface is crucial for Breedbase. Breedbase uses BrAPI to connect with the data collection apps, other projects such as CLIMMOB [6], and native BrAPPs built into the Breedbase webpage. Users also appreciate the ability to connect to Breedbase instances using packages such as QBMS [46] for data import into R for custom analyses. The Breedbase team has been part of the BrAPI community since its inception and has continuously adopted and contributed to the BrAPI standard.

## DeltaBreed

DeltaBreed is an open-source breeding data management system designed and developed by Breeding Insight to support U.S. Department of Agriculture - Agricultural Research Service (USDA-ARS) specialty crop and animal breeders. DeltaBreed differs from other related systems in that it is customizable to small breeding teams and generalized enough to support the workflows of diverse species. DeltaBreed is a unified system that connects a variety of BrAPI applications. BrAPI integration allows the complexity underlying interoperability to be hidden, shielding users from multifactorial differences between various applications. DeltaBreed, adhering to the BrAPI model, establishes data standards and validations for users and provides a singular framework for data management and user training. BrAPI enabled connections are being used with all of the following tools: BrAPI Java Test Server, BreedBase, Field Book, Gigwa, QBMS, Mr Bean, Helium and the Pedigree Viewer BrAPP.

DeltaBreed users may not be aware of BrAPI or the specifics of underlying tools, but will notice that BrAPI interoperability reduces the need for human-mediated file transfers and data manipulation. Field Book users, for example, can connect to their DeltaBreed program, authenticate, and pull studies and observation variables directly from DeltaBreed to Field Book on their data collection device. The subsequent step of pushing observations from Field Book to DeltaBreed is straightforward via BrAPI, but is pending implementation until data quality validations are put in place; these include improved data transaction handling and differentiation of intentional and inadvertent repeated measures.

## FAIDARE

FAIDARE [47] is a data discovery portal providing a biologist-friendly search system over a global federation of 40 plant research databases. It allows users to identify data resources using a full text search approach combined with domain specific filters. Each search result contains a link back to the original database for visualization, analysis, and download. The indexed data types are broad and include genomic features, selected bibliography, QTL, markers, genetic variation studies, phenomic studies, and plant genetic resources. This inclusiveness is achieved thanks to a two stage indexation data model. The first index, more generic, provides basic search functionalities and relies on five fields: name, link back URL, data type, species, and exhaustive description. To provide more advanced filtering, the second stage indexation mechanism takes advantage of BrAPI endpoints to get more detailed metadata on germplasm, genotyping studies and phenotyping studies.

The FAIDARE indexation mechanism relies on a public software package [48] that allows data resource managers to request the indexation of their database. This BrAPI client is currently able to extract data from any BrAPI v1.3 and v1.2 endpoint, and the development of BrAPI v2.x indexation will be initiated in 2025. Since not all databases are willing to implement BrAPI endpoints, it is possible to generate metadata as static BrAPI-compliant JSON files, using the BrAPI standard as a file exchange format.

The FAIDARE architecture has been designed by elaborating on the BrAPI data model in combination with the GnpIS Software Architecture [32]. It uses an Elasticsearch NoSQL engine that searches and serves enriched versions of the BrAPI JSON data model. FAIDARE also includes a BrAPI endpoint using all indexed metadata. It has been adopted by several communities including the ELIXIR and EMPHASIS European infrastructures, and the WheatIS of the Wheat-Initiative. Several databases are added each year to the FAIDARE global federation, adding to both the portal and BrAPI adoption.

### Germinate

Germinate [49,50] is an open-source plant genetic resources database that combines and integrates various types of plant breeding data including genotypic, phenotypic, passport, image, geographic, and climate data into a single repository. Germinate is tightly linked to the BrAPI specification and supports the majority of BrAPI endpoints for querying, filtering, and submission.

Germinate connects with other BrAPI-enabled tools such as GridScore for phenotypic data collection, Flapjack for genotypic data visualization, and Helium for pedigree visualization. Additionally, due to the nature of BrAPI, Germinate can act as a data repository for any BrAPI-compatible tool. The interoperability provided by BrAPI reduces the need for manual data handling, providing the direct benefits of faster data processing, fewer human errors, and improved data security and integrity.

## Analytics

Modern breeding programs have multiple decision points requiring analysis and integration of various data types. While there are numerous Breeding and Genetics Data Management systems (above), certain program tasks could be simplified by development of specific streamlined analysis systems. These analysis systems better enable certain tasks by utilizing data from different sources to make efficient data-driven decisions. With increased computational power at their disposal, scientists can construct more advanced analysis pipelines by combining various data sources.

The tools developed by the BrAPI community can pull in data from multiple BrAPI-compatible data sources and provide enhanced analytical functionality. In many cases, there is no longer a need to import and export large data files to a local computational environment just to run standard analytical models. These tools are able to extract the data they need from various data sources without much

human intervention or human error. They can also provide simple user interaction to enhance decision support for the breeders and researchers.

## G-Crunch

G-Crunch is an upcoming user-facing tool to make simple, repeatable analysis requests. The lightweight UI can be used to specify and filter incoming data, select specific analysis criteria, and trigger any analytics pipeline that is built into the specific framework instance. G-Crunch is currently built on top of the open-source Analytics Framework project, and can run pipelines using tools such as sommer and ASREML. Each piece of the data and pipeline can be separately specified, which can allow flexibility when running complex analysis. A 'test' analysis can be run on small data sets with small or local analytics engine, then quickly re-direct G-Crunch to a larger dataset and a larger computational framework. This mitigates the complications of moving data around and introducing errors from manually triggering the analysis steps.

G-Crunch relies on BrAPI endpoints to access phenotypic and genotypic data sources, as well as an API currently implemented in the Analytics Framework to start and track processes. G-Crunch, as a tool, couldn't feasibly exist without BrAPI. The support of BrAPI interfaces allows G-Crunch to use one unified request method and adapt to the user's existing network of BrAPI-compliant tools. This lowers the barrier to entry for adoption and makes analysis pipelines easily repeatable.

## QBMS

Many plant breeders and geneticists analyze their datasets using the R statistical programming language, but this requires the import of data into an R environment. BrAPI enables access to pull datasets into R from compatible databases, but API backend processes, such as authentication, tokens, TCP/IP protocol, JSON format, pagination, stateless calls, asynchronous communication, and database IDs are complex for users to navigate. The QBMS R package eliminates technical barriers scientists experience when using the BrAPI specification in their analysis scripts and pipelines by providing breeders with stateful functions familiar to them when navigating their GUI systems [46]. QBMS enables users to query and extract data into a dataframe, a common structure in the R language, providing an intuitive connection with breeding data management systems.

The community has built extended solutions on top of QBMS, incorporating the package into R-Shiny BrAPPs such as Mr.Bean [51] (described below). QBMS is open-source and available on the official CRAN repository, where it has garnered over 9400 downloads.

## Mr.Bean

Mr.Bean [51] is a graphical user interface (GUI) designed to assist breeders, statisticians, and individuals involved in plant breeding programs with the analysis of field trials. By utilizing innovative methodologies such as SpATS for modeling spatial trends, and autocorrelation models to address spatial variability. Mr.Bean proves highly practical and powerful in facilitating faster and more effective decision-making. Modeling Genotype-by-environment interaction poses its challenges, but Mr.Bean offers the capability to explore various variance-covariance matrices, including Factor Analytic, compound symmetry, and heterogeneous variances. This aids in the assessment of genotype performance across diverse environments.

Mr.Bean boasts flexibility in importing different file types, yet for users managing their data within data management systems, the process of downloading from their systems and importing it into Mr.Bean can be cumbersome. To address this issue, QBMS was integrated into the back end. This feature prompts users to input the URL of a BrAPI compatible server, enter their credentials (if

necessary), and select the specific trial they wish to analyze. Subsequently, users can seamlessly access their dataset through BrAPI and utilize it across the entire Mr.Bean interface.

## SCT

The Sugarcane Crossing Tool (SCT) is a lightweight R-Shiny dashboard application designed to receive, process, and visualize data from a linked BreedBase [13] instance. This application is being developed collaboratively with members of the Sugarcane Integrated Breeding System, who have advocated for an application that assists them in designing crosses based on queried information from a list of available accessions. By leveraging existing community resources, the team has been able to develop a simple, BrAPI-enabled, application without possessing extensive programming knowledge or experience. The SCT is presented as an inspiration for similarly positioned scientists to consider developing custom applications for specific tasks.

The crossing tool utilizes a modified version of the BrAPI-R library to access a compliant database, and it employs standard R/JavaScript packages to aggregate and visualize data. Modules within the application allow breeders to query the database (through BrAPI) for information relevant to their decision-making process. This includes the number and sex of flowering accessions, deep pedigree and relatedness information, summarized trial data, and the prior frequency and success of potential cross combinations. Future versions of this tool will provide additional decision support (e.g. ranked potential crosses) to enhance the accuracy and efficiency of crossing.

## ShinyBrAPPs

The ShinyBrAPPs code repository contains a number of useful tools, built using the R-Shiny framework and the BrAPI-R open-source library. The R-Shiny framework allows user communities to quickly prototype and produce applications that are finely tailored to their needs, thus improving adoption and daily use of data management systems. An international collaboration of developers from CIRAD and the IBP have been working together as part of the IAVAO breeders community to develop these ShinyBrAPPs, in support of national breeding programs in western Africa. These applications are typically connected to BMS and/or Gigwa and provide tools for specific use cases. BrAPI compliance offers these systems the opportunity to add functionalities in a modular way through the development of external plugin applications that can quickly fulfill specific needs for this group of breeders and scientists.

So far, four applications have been developed covering the fields of trial data quality control, single trial statistical analysis, breeding decision support, and raw genotyping data visual inspection. The "BMS trial data explorer" retrieves data from a single multi-location trial and displays data counts and summary box-plots for all variables measured in different studies. It also provides an interactive distribution plot to easily select observations that require curation and a report of candidate issues that needs to be addressed by the breeder. The "STABrAPP" tool is an application for single trial mixed model analysis. It basically provides a GUI to the StatGen-STA R package. The "DSBrAPP" tool is a decision support tool helping breeders to select germplasm according to their various characteristics and save this germplasm list into BMS. Finally, the "snpclust" tool enables a user to check and manually correct the clustering of fluorescence based SNP genotyping data.

## General Infrastructure

Adopting BrAPI compatibility into an existing system can be difficult sometimes. The BrAPI Community has developed several tools to make adoption easier, this set of tools are built to support the programmers and developers. This includes things like prebuilt code libraries, connectors to other technology standards, and mappers to alternate data types or data files. The goal is to lower the

barrier to entry for the BrAPI community, making it easier for other groups to get started and connect their existing data to the standard.

## BrAPIMapper

BrAPIMapper is a full BrAPI implementation designed to be a convenient wrapper for any breeding related data source. BrAPIMapper is provided as a Docker application that can connect to a variety of external data sources including mySQL or PostgreSQL databases, generic REST services, flat files (XML, JSON, CSV/TSV/GFF3/VCF, YAML), or any combination of these. It provides an administration user interface to map BrAPI data models to external data sources. The interface allows administrators to select the BrAPI specification versions to use and which endpoints to enable. Data mapping configuration import and export features simplify upgrades to future BrAPI versions; administrators only have to map missing fields or make minor adjustments. BrAPIMapper supports the primary BrAPI features including paging, deferred search results, user lists, and authentication. Access restrictions to specific endpoints can be managed through the administration interface as well. This tool aims to accelerate BrAPI services deployment while ensuring specification compliance.

## MIRA and BrAPI2ISA

Since the release of BrAPI 1.3, efforts have been made to incorporate support for the MIAPPE (Minimal Information About a Plant Phenotyping Experiment) [10] standard into the specification, achieving full compatibility in BrAPI 2.0. Consequently, BrAPI now includes all attributes necessary for MIAPPE compliance, adhering to standardized descriptions in accordance with MIAPPE guidelines. In some communities and projects, phenotyping data and metadata are archived and published as structured ISA-Tab files, validated using the MIAPPE ISA configuration [52]. Although ISA-Tab is easy to read for non-technical experts due to its file-based approach, it lacks programmatic accessibility, particularly for web applications.

MIRA enables the automatic deployment of a BrAPI server on a MIAPPE-compliant dataset in ISA-Tab format, facilitating programmatic access to these datasets. It is deployable from a Docker image with the dataset mounted. The tool leverages the mapping between MIAPPE, ISA-Tab, and BrAPI, eliminating the need for parsing or manual mapping of datasets compliant with (meta-)data standards. By providing programmatic access through BrAPI, MIRA facilitates the integration of phenotyping datasets into web applications.

The BrAPI2ISA service functions as a converter between a BrAPI-compatible server and the ISA-Tab format. The tool simplifies, automates, and facilitates the archiving of data, thereby enhancing data preservation and accessibility. The BrAPI2ISA tool is compatible with BrAPI 1.3 and welcomes community contributions to support the latest versions of BrAPI.

## GraphQL Data-warehouse

Using the Zendro set of automatic software code generators, a fully functional, efficient, and cloud-capable BrAPI data-warehouse has been created for the current version of the BrAPI data models. Unlike most BrAPI-compliant data sources, this data-warehouse supports a GraphQL API rather than a RESTful API. This API provides secure access to data read and write functions for all BrAPI data models. It provides create, read, update, and delete (CRUD) functions that are standardized and accept the same parameters for all data models. Zendro supports a large number of underlying database systems, allowing flexibility during installation and integration.

The GraphQL server is particularly rich in features. Logical filters allow for exhaustive search queries, whose structure is highly intuitive and based around logical triplets. Such triplets consists of a BrAPI

model property, a logical operator, and a value, e.g. "Study-name equals 'Nursery Study'". A large collection of operators is available and triplets can be combined to logical search trees using "and" or "or" operators. Searches can be extended over relationships between data models, thus enabling a user to query the warehouse for exactly the required data. Authorization is based on user roles and can be configured differently for each single data model read or write function. The generated graphical interface allows for the integration of interactive scientific plots and analysis tools written in JavaScript or WebAssembly.

An example data warehouse is publicly available and offers full read access in the graphical user interface and through the GraphQL API. The example warehouse is populated with public CassavaBase data [53] to create fully BrAPI-compliant example based on Zendro. Three interactive scientific example plots are available to explore the data. The first is a boxplot comparing Cassava harvest indices measured for four different experiments. Next, an interactive raincloud plot provides an alternative visualization of the same data. Finally, a scatterplot shows how Cassava fresh root yield and plant height are correlated based on data from a single study.

# Discussion

## BrAPI for Breeders

While the BrAPI technical specification is designed to be read and used by software developers, its underlying purpose is to support the work of breeders and other scientists by making routine processes faster, easier, and cheaper. BrAPI offers a convenient path to automation, interoperability, and data integration for software tools in breeding, genetics, phenomics, and other related agricultural domains. By integrating the tools described above, breeders and scientists can spend less time on data management and more time focusing on science. For many, the ultimate goal is the development of a digital data ecosystem: a collection of software tools and applications that can all work together seamlessly. In this scenario, data is digitally collected, automatically sent to quality control systems, batch analyzed to provide actionable insights, and finally stored in accessible databases for long-term applications. As tools continue to adopt the BrAPI standard, this vision is beginning to approach reality.

## Looking Ahead

The BrAPI project leadership and community are committed to building standards to support new use cases and technologies as they are adopted by breeders and other scientists, potentially including drone imaging data, spectroscopy, LIDAR, metabolomics, transcriptomics, agronomics, high-throughput phenotyping, pangenomics, and machine learning based analysis. Each of these technologies will have unique challenges, generate different types of data, and require substantial thought and discussion before being added to the BrAPI specification. This process has already begun for several data types, with small groups working to build generic data models and proposed communication standards. As these community efforts are completed, they will become part of a future version of the BrAPI standard, enabling further interoperability and simplifying data exchange.

Expanding the BrAPI specification is important for the community, but this growth should not reinvent or compete with existing functional standards. Additions to the BrAPI specification are reviewed thoroughly by the community to make sure BrAPI is compliant with existing standards and data structures. For example, the community has requested compliance with the GFF3 standard for genomic data and the GeoTIFF standard for aerial image data. Pieces of these existing popular data structures might be integrated into the overall BrAPI standard documentation. In some cases, BrAPI will only reference other standards instead of including them in the specification. For example, there have been community discussions around developing connections with the NOAA CDO standard for

weather data or the Galaxy Analytics API for analytics pipeline controls and information. These standards are perfectly adequate on their own and recreating them in the BrAPI standard would be redundant.

## Conclusion

The BrAPI project only exists because of the community of software engineers, biologists, and other scientists who support and use it. While there were many tools and use cases presented here, it is not an exhaustive list of all BrAPI-compliant systems. As long as the standard continues to be supported, the community will continue to expand. As more groups continue to make their tools BrAPI compliant, others will see the value in implementing BrAPI into their own tools, allowing the community to strengthen and grow. By providing an open standard for breeding data, and the infrastructure and community to support it, the BrAPI project is doing its part to support a productive agricultural system amidst the pressing challenges of climate change. If this manuscript is your first introduction to the BrAPI project, the authors invite you to join the community. More information is available at brapi.org.

## Methods

The BrAPI Project Coordinator is responsible for the day-to-day operations and general maintenance of the project infrastructure, as well as coordinating updates to the standard. They organize community events and encourage collaboration between community groups. Long term planning and organization is handled by the BrAPI Advisory board. This is a board of six community members who are elected by the community to represent their interest in the project.

The standard documentation is stored and maintained in a public GitHub repository. The core documentation is written using the OpenAPI 3 documentation standard. The core documentation is automatically published on Apiary and SwaggerHub, two API documentation hosting sites. The specification has a standard MIT open-source license.

New versions of the BrAPI standard are developed based on community demand. Opportunities to enhance and improve the standard are identified by the community, tracked in GitHub issues, and implemented to create a new stable version. This process ensures the standard stays relevant, updated, and stable over time. Minor version updates are designed to be backwards compatible.

Hackathon events are the primary approach used by the BrAPI community to foster collaboration and further development. One in-person and one virtual event per year maintain project momentum and social comradery. The hackathons provide a dedicated time to discuss improvements and issues with the specification, particular use cases, and project stewardship.

## Acknowledgements

# References

1.  Kuriakose SV, Pushker R, Hyde EM. Data-Driven Decisions for Accelerated Plant Breeding. Accelerated Plant Breeding, Volume 1. Springer International Publishing; 2020. pp. 89–119. doi:10.1007/978-3-030-41866-3_4

2.  Selby P, Abbeloos R, Backlund JE, Basterrechea Salido M, Bauchet G, Benites-Alfaro OE, et al. BrAPI—an application programming interface for plant breeding applications. Wren J, editor. Bioinformatics. 2019;35: 4147–4155. doi:10.1093/bioinformatics/btz190

3.  Ghouila A, Siwo GH, Entfellner J-BD, Panji S, Button-Simons KA, Davis SZ, et al. Hackathons as a means of accelerating scientific discoveries and knowledge transfer. Genome Res. 2018;28: 759–765. doi:10.1101/gr.228460.117

4.  Rife TW, Poland JA. Field Book: An Open-Source Application for Field Data Collection on Android. Crop Science. 2014;54: 1624–1627. doi:10.2135/cropsci2013.08.0579

5.  Raubach S, Schreiber M, Shaw PD. GridScore: a tool for accurate, cross-platform phenotypic data collection and visualization. BMC Bioinformatics. 2022;23. doi:10.1186/s12859-022-04755-2

6.  Quirós C, de Sousa K, Steinke J, Madriz B, Laporte M-A, Arnaud E, et al. ClimMob: Software to support experimental citizen science in agriculture. Computers and Electronics in Agriculture. 2024;217: 108539. doi:10.1016/j.compag.2023.108539

7.  Morales N, Kaczmar NS, Santantonio N, Gore MA, Mueller LA, Robbins KR. ImageBreed: Open-access plant breeding web–database for image-based phenotyping. The Plant Phenome Journal. 2020;3. doi:10.1002/ppj2.20004

8.  Neveu P, Tireau A, Hilgert N, Nègre V, Mineau-Cesari J, Brichet N, et al. Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. New Phytologist. 2018;221: 588–601. doi:10.1111/nph.15385

9.  Available: https://urgi.versailles.inrae.fr/faidare

10. Papoutsoglou EA, Faria D, Arend D, Arnaud E, Athanasiadis IN, Chaves I, et al. Enabling reusability of plant phenomic datasets with MIAPPE 1.1. New Phytologist. 2020;227: 260–273. doi:10.1111/nph.16544

11. PIPPA - PSB Interface for Plant Phenotype Analysis. [cited 25 Sep 2024]. Available: https://pippa.psb.ugent.be/

12. Available: https://www.wiwam.be

13. Morales N, Ogbonna AC, Ellerbrock BJ, Bauchet GJ, Tantikanjana T, Tecle IY, et al. Breedbase: a digital ecosystem for modern plant breeding. Morris G, editor. G3 Genes|Genomes|Genetics. 2022;12. doi:10.1093/g3journal/jkac078

14. Ruas M, Guignon V, Sempere G, Sardos J, Hueber Y, Duvergey H, et al. MGIS: managing banana (Musa spp.) genetic resources information and high-throughput genotyping data. Database. 2017;2017. doi:10.1093/database/bax046

15. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158. doi:10.1093/bioinformatics/btr330

16. ga4gh-metadata/SchemaBlocks. GA4GH Metadata Schema Development Team; 2023. Available: https://github.com/ga4gh-metadata/SchemaBlocks

17. König P, Beier S, Mascher M, Stein N, Lange M, Scholz U. DivBrowse—interactive visualization and exploratory data analysis of variant call matrices. GigaScience. 2022;12. doi:10.1093/gigascience/giad025

18. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10. doi:10.1186/1471-2105-10-421

19. Cock PJA, Chilton JM, Grüning B, Johnson JE, Soranzo N. NCBI BLAST+ integrated into Galaxy. GigaSci. 2015;4. doi:10.1186/s13742-015-0080-7

20. Milne I, Shaw P, Stephen G, Bayer M, Cardle L, Thomas WTB, et al. Flapjack—graphical genotype visualization. Bioinformatics. 2010;26: 3133–3134. doi:10.1093/bioinformatics/btq580

21. Sempéré G, Pétel A, Rouard M, Frouin J, Hueber Y, De Bellis F, et al. Gigwa v2—Extended and improved genotype investigator. GigaScience. 2019;8. doi:10.1093/gigascience/giz051

22. Rouard M, Sardos J, Sempéré G, Breton C, Guignon V, Van den Houwe I, et al. A digital catalog of high-density markers for banana germplasm collections. Plants People Planet. 2021;4: 61–67. doi:10.1002/ppp3.10187

23. Bradbury PJ, Casstevens T, Jensen SE, Johnson LC, Miller ZR, Monier B, et al. The Practical Haplotype Graph, a platform for storing and using pangenomes for imputation. Kendziorski C, editor. Bioinformatics. 2022;38: 3698–3702. doi:10.1093/bioinformatics/btac410

24. Engels JMM, Ebert AW. A Critical Review of the Current Global Ex Situ Conservation System for Plant Agrobiodiversity. I. History of the Development of the Global System in the Context of the Political/Legal Framework and Its Major Conservation Components. Plants. 2021;10: 1557. doi:10.3390/plants10081557

25. Fu Y. The Vulnerability of Plant Genetic Resources Conserved Ex Situ. Crop Science. 2017;57: 2314–2328. doi:10.2135/cropsci2017.01.0014

26. Global Information System. [cited 25 Sep 2024]. Available: https://glis.fao.org/glis/

27. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3. doi:10.1038/sdata.2016.18

28. Beier S, Fiebig A, Pommier C, Liyanage I, Lange M, Kersey PJ, et al. Recommendations for the formatting of Variant Call Format (VCF) files to make plant genotyping data FAIR. F1000Res. 2022;11: 231. doi:10.12688/f1000research.109080.2

29. Adam-Blondon A-F, Alaux M, Brouwer M, Kersey P, Lange M, Le Floch E, et al. AGENT Guidelines for dataflow. Zenodo; 2024. doi:10.5281/zenodo.12625360

30. Street K, Street K. Genebank mining with FIGS, the Focused Identification of Germplasm Strategy. Unknown. 2017. doi:10.22004/ag.econ.266624

31. Kotni P, van Hintum T, Maggioni L, Oppermann M, Weise S. EURISCO update 2023: the European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material. Nucleic Acids Research. 2022;51: D1465–D1469. doi:10.1093/nar/gkac852

32. Pommier C, Michotey C, Cornut G, Roumet P, Duchêne E, Flores R, et al. Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. Plant Phenomics. 2019;2019.

doi:[10.34133/2019/1671403](https://doi.org/10.34133/2019/1671403)

33. Adam-Blondon A-F, Alaux M, Durand S, Letellier T, Merceron G, Mohellibi N, et al. Mining Plant Genomic and Genetic Data Using the GnpIS Information System. Methods in Molecular Biology. Springer New York; 2016. pp. 103–117. doi:[10.1007/978-1-4939-6658-5_5](https://doi.org/10.1007/978-1-4939-6658-5_5)

34. Shaw PD, Graham M, Kennedy J, Milne I, Marshall DF. Helium: visualization of large scale plant pedigrees. BMC Bioinformatics. 2014;15. doi:[10.1186/1471-2105-15-259](https://doi.org/10.1186/1471-2105-15-259)

35. Paterson T, Law A. G<scp>enotypechecker</scp>: an interactive tool for checking the inheritance consistency of genotyped pedigrees. Animal Genetics. 2011;42: 560–562. doi:[10.1111/j.1365-2052.2011.02183.x](https://doi.org/10.1111/j.1365-2052.2011.02183.x)

36. Bernád V, Al-Tamimi N, Langan P, Gillespie G, Dempsey T, Henchy J, et al. Unlocking the genetic diversity and population structure of the newly introduced two-row spring European HerItage Barley collecTion (ExHIBiT). Front Plant Sci. 2024;15. doi:[10.3389/fpls.2024.1268847](https://doi.org/10.3389/fpls.2024.1268847)

37. Elsner-Gearing F, Kretzschmar P, Shultz S, Pilgrim M, Dawson DA, Horsburgh GJ, et al. Admixture and reproductive skew shape the conservation value of ex situ populations of the Critically Endangered eastern black rhino. Conserv Genet. 2024;25: 897–910. doi:[10.1007/s10592-024-01611-z](https://doi.org/10.1007/s10592-024-01611-z)

38. Van den houwe I, Chase R, Sardos J, Ruas M, Kempenaers E, Guignon V, et al. Safeguarding and using global banana diversity: a holistic approach. CABI Agric Biosci. 2020;1. doi:[10.1186/s43170-020-00015-6](https://doi.org/10.1186/s43170-020-00015-6)

39. Jung S, Lee T, Gasic K, Campbell BT, Yu J, Humann J, et al. The Breeding Information Management System (BIMS): an online resource for crop breeding. Database. 2021;2021. doi:[10.1093/database/baab054](https://doi.org/10.1093/database/baab054)

40. Shrestha R, Arnaud E, Mauleon R, Senger M, Davenport GF, Hancock D, et al. Multifunctional crop trait ontology for breeders' data: field book, annotation, data discovery and semantic enrichment of the literature. AoB PLANTS. 2010;2010. doi:[10.1093/aobpla/plq008](https://doi.org/10.1093/aobpla/plq008)

41. Jung S, Lee T, Cheng C-H, Buble K, Zheng P, Yu J, et al. 15 years of GDR: New data and functionality in the Genome Database for Rosaceae. Nucleic Acids Research. 2018;47: D1137–D1145. doi:[10.1093/nar/gky1000](https://doi.org/10.1093/nar/gky1000)

42. Yu J, Jung S, Cheng C-H, Lee T, Zheng P, Buble K, et al. CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research. Plants. 2021;10: 2805. doi:[10.3390/plants10122805](https://doi.org/10.3390/plants10122805)

43. Home. In: Breeder Genomics Hub [Internet]. [cited 25 Sep 2024]. Available: [https://hub.maizegenetics.net/](https://hub.maizegenetics.net/)

44. Morales N, Bauchet GJ, Tantikanjana T, Powell AF, Ellerbrock BJ, Tecle IY, et al. High density genotype storage for plant breeding in the Chado schema of Breedbase. Wang J, editor. PLoS ONE. 2020;15: e0240059. doi:[10.1371/journal.pone.0240059](https://doi.org/10.1371/journal.pone.0240059)

45. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics. 2001;157: 1819–1829. doi:[10.1093/genetics/157.4.1819](https://doi.org/10.1093/genetics/157.4.1819)

46. Khaled Al-Shamaa, Johan Steven Aparicio, Nick, icarda-git. icarda-git/QBMS: QBMS Version 1.0.0. Zenodo; 2024. doi:[10.5281/zenodo.10791627](https://doi.org/10.5281/zenodo.10791627)

47. Pommier C. FAIR Plant Phenomics Data Management Tools and Guidelines. 2024. Available: [https://hal.science/hal-04425516](https://hal.science/hal-04425516)

48. elixir-europe/plant-brapi-etl-faidare. ELIXIR Europe; 2024. Available: https://github.com/elixir-europe/plant-brapi-etl-faidare

49. Raubach S, Kilian B, Dreher K, Amri A, Bassi FM, Boukar O, et al. From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives. Crop Science. 2020;61: 1538–1566. doi:10.1002/csc2.20248

50. Shaw PD, Raubach S, Hearne SJ, Dreher K, Bryan G, McKenzie G, et al. Germinate 3: Development of a Common Platform to Support the Distribution of Experimental Data on Crop Wild Relatives. Crop Science. 2017;57: 1259–1273. doi:10.2135/cropsci2016.09.0814

51. Aparicio J, Gezan SA, Ariza-Suarez D, Raatz B, Diaz S, Heilman-Morales A, et al. Mr.Bean: a comprehensive statistical and visualization application for modeling agricultural field trials data. Front Plant Sci. 2024;14. doi:10.3389/fpls.2023.1290078

52. Sansone S-A, Rocca-Serra P, Field D, Maguire E, Taylor C, Hofmann O, et al. Toward interoperable bioscience data. Nat Genet. 2012;44: 121–126. doi:10.1038/ng.1054

53. Fernandez-Pozo N, Menda N, Edwards JD, Saha S, Tecle IY, Strickler SR, et al. The Sol Genomics Network (SGN)—from genotype to phenotype to breeding. Nucleic Acids Research. 2014;43: D1036–D1041. doi:10.1093/nar/gku1195