BrAPI Success Stories

This manuscript (permalink) was automatically generated from plantbreeding/BrAPI-Manuscript2@ca78d90 on April 17, 2024.

Authors

- **Peter "BrapMan" Selby** ✉
  0000-0001-7151-4445 · BrapiCoordinatorSelby
  Cornell University · Funded by NIFA-DSFAS 2022-67021-37024

- **Trevor "Cool Kid" Rife**
  0000-0002-5974-6523 · trife
  Clemson University

- **Khaled Al-Shamaa**
  0000-0002-7668-3798 · khaled-alshamaa
  ICARDA

- **Isabelle Alic**
  0000-0002-8961-6068 · Isabelle-inrae
  INRAE · Funded by {'Short version': 'This work was supported by the project Phenome-ANR-11-INBS-0012'}; {'OR Long version': "This work was supported by the Agence Nationale de la Recherche, programme Investissements d'avenir, ANR-11-INBS-0012 (Phenome)"}

- **Sebastian "Baz" Raubach**
  0000-0001-5659-247X · sebastian-raubach
  The James Hutton Institute

- **Paul Shaw**
  0000-0002-0202-1150 · cardinalb
  The James Hutton Institute

- **Iain Milne**
  0000-0002-4126-0859 · imilne
  The James Hutton Institute

- **Cyril Pommier**
  0000-0002-9040-8733 · cpommier
  Université Paris-Saclay, INRAE, BioinfOmics, Plant Bioinformatics Facility, Versailles, France; Université Paris-Saclay, INRAE, URGI, Versailles, France · Funded by {'Short version': 'This work was supported by the project Phenome-ANR-11-INBS-0012'}; {'OR Long version': "This work was supported by the Agence Nationale de la Recherche, programme Investissements d'avenir, ANR-11-INBS-0012 (Phenome)"}

- **Becky Smith**
  0000-0002-8968-3383 · Batbaby91
  The James Hutton Institute

- **Chaney Courtney**
  · chaneylc
  Clemson University

- **Zachary Miller**
  0000-0002-5454-4527 · zrm22
  Cornell University

- **Terry Casstevens**
  0000-0001-7602-0487 · tcasstevens
  Cornell University

- **Sook Jung**
  0000-0003-3968-2769
  Department of Horticulture, Washington State University

- **Taein Lee**
  0000-0002-2000-2586 · leetaei
  Department of Horticulture, Washington State University

- **Josh "Big Data" Lamos-Sweeney**
  · jlamossweeney
  Cornell University

- **Francisco López**
  International Treaty on Plant Genetic Resources for Food and Agriculture, FAO

- **Marco Marsella**
  0000-0003-0334-8785
  International Treaty on Plant Genetic Resources for Food and Agriculture, FAO

- **Matthias Lange**
  0000-0002-4316-078X · langeipk
  Leibniz Institute of Plant Genetics and Crop Plant Research

- **Guilhem Sempéré**
  0000-0001-7429-2091 · GuilhemSempere
  CIRAD (french agricultural research and international cooperation organization); South Green Platform

- **Stephan Weise**
  Leibniz Institute of Plant Genetics and Crop Plant Research

- **Patrick König**
  0000-0002-8948-6793 · patrick-koenig
  Leibniz Institute of Plant Genetics and Crop Plant Research

- **Gouripriya Davuluri**
  Leibniz Institute of Plant Genetics and Crop Plant Research

- **Paul Kersey**
  Royal Botanic Gardens, Kew

- **Erwan Le-Floch**
  URGI PlantBioinfoPF, INRAE France

- **Jospeh Ruff**
  Royal Botanic Gardens, Kew

- **Michael Alaux**
  URGI PlantBioinfoPF, INRAE France

- **Suman Kumar**
  Leibniz Institute of Plant Genetics and Crop Plant Research

- **Matthijs Brouwer**
  Wageningen University and Research

- **Bert Droesbeke**
  0000-0003-0522-5674 · bedroesb
  VIB Data Core

- **Valentin Guignon**
  0000-0003-0903-6811 · guignonv
  The Alliance of Bioversity International and CIAT (CGIAR)

✉ — Correspondence possible via GitHub Issues or email to Peter "BrapMan" Selby <ps664@cornell.edu>.

Abstract

The Breeding API (BrAPI) project is an effort to enable interoperability among plant breeding databases. BrAPI is a standardized RESTful web service API specification for communicating plant breeding data. This community driven standard is free to be used by anyone interested in plant breeding data management. This manuscript describes updates and outlook for the current version of BrAPI.

Introduction

Plant and animal breeding is an incredibly important part of today's society. Almost every country in the world has some kind of breeding program supporting the agricultural community to produce bigger, better, healthier, more sustainable crops. Modern breeding techniques require large amounts of high quality data to be effective. In the digital age, that breeding data is being collected, managed, and analyzed with computer software. Interoperability between breeding software tools, systems, and databases can substantially increase the efficiency of a breeding program. The ability to share tools gives each program a boost in computational power. The ability to share data means everyone has access to larger, more complete, datasets and get build more accurate computational models and produce more accurate predictions.

The Breeding API (BrAPI) project is an effort to enable interoperability among breeding tools, systems, and databases. BrAPI is a standardized Representational State Transfer (REST), web service, Application Programming Interface (API), specification for breeding and related agricultural data. [1] By using the BrAPI standard, breeding software can more easily become interoperable, allowing groups to more easily share data and software tools.

How it works

An Application Programming Interface (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 GUI or CLI might allow a user to input data, read data, and start processes within an application. An API allows one piece of software (sometimes called a client, user agent, or service consumer) to programmatically input data, read data, and start process within another piece of software (sometimes called a server or service provider).

A Representational State Transfer (REST) web service is a type of API commonly used in today's modern web infrastructure. REST is a technical architecture that describes the stateless transmission of data between applications. Typically, REST systems are implemented using the standard HTTP protocol that most of the modern internet is built upon. REST implementations also generally use JavaScript Object Notation (JSON) to represent the data being transferred. Both HTTP and JSON are programming language agnostic, very stable, and very 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 can choose to represent their data as a BrAPI compatible API. By mapping the internal data structures to the standard models, data repositories can easily expose data to the outside world. Similarly, they can accept new data from external sources and automatically map the new data into the existing database. Client application developers can take advantage of this standardization by building tools that can easily integrate with all other BrAPI compatible data repositories. Visualization, reporting, analytics, data collection, and quality control tools can be built once and shared with other organizations following the standards. As the number of BrAPI compatible databases, tools, and organizations grows, so does the value added by implementing the standard into a given application.

Project Updates

Over its lifetime, the BrAPI project has grown and changed substantially. The latest stable version of the specification (v2.1) looks vastly different from the original version (v1.0) released in 2017. The total size of the specification has almost quadrupled in that time, going from 51 endpoints documented in v1.0 to 201 endpoints documented in v2.1. Because of this growth, the specification documents were reorganized into four modules: BrAPI-Core, BrAPI-Germplasm, BrAPI-Genotyping, and BrAPI-Phenotyping. Figure 1 shows a simplified domain map of the whole BrAPI v2.1 data model, divided into the organizational modules. The early versions of the specification focused on read-only phenotype data, with a small consideration to the other domains. Now the specification has a full representation of most of the major concepts applicable to the breeding process. The new specification is also internally consistent, easier to navigate, and allows for read, write, and update capabilities. None of those qualities were a guarantee for the earlier versions.

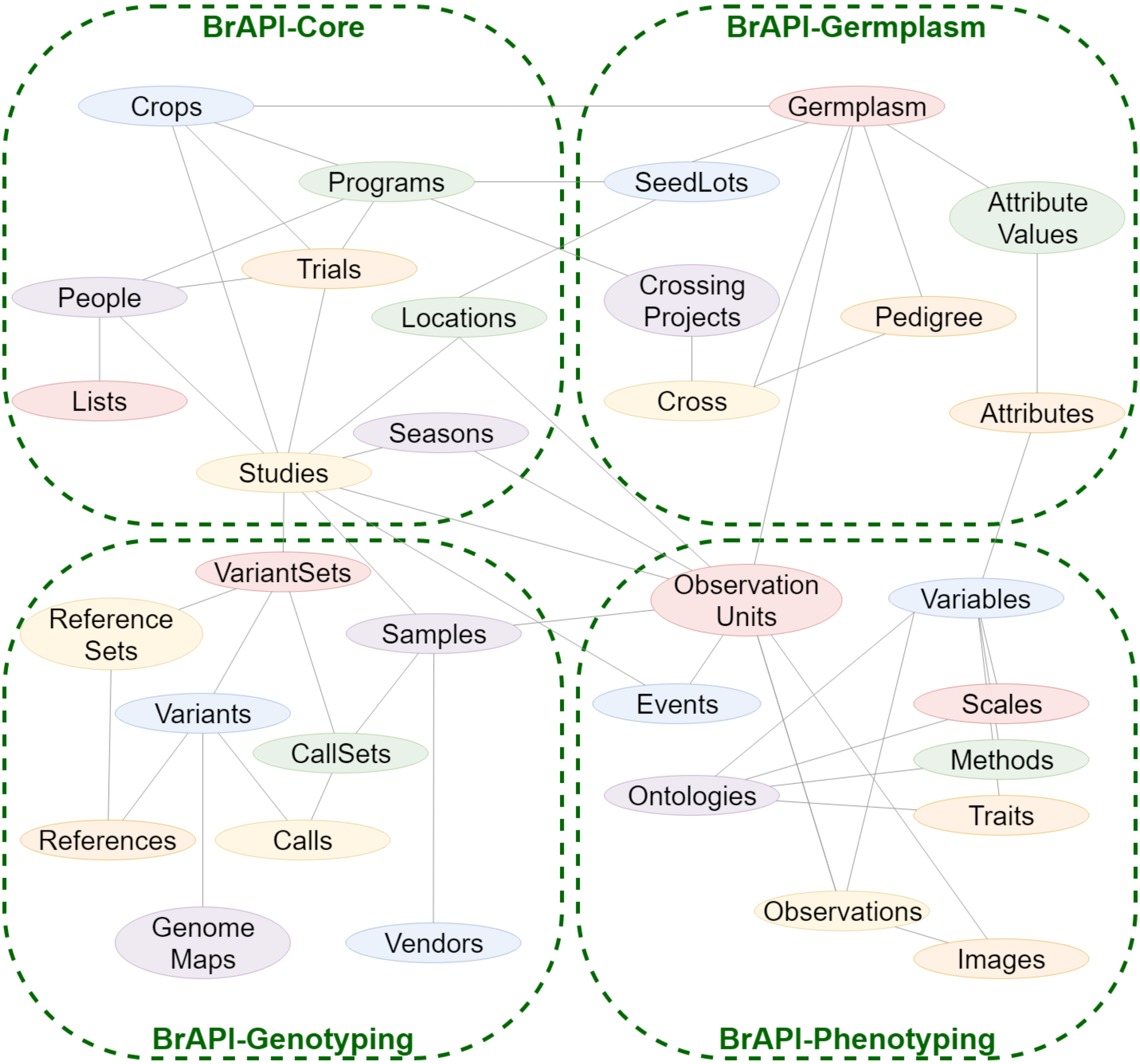

Figure 1: A simplified domain map of the whole BrAPI data model, divided into organizational modules. A more detailed Entity Relationship Diagram (ERD) is available on brapi.org.

As the specification has matured, so have the tools, services, and libraries available to the community to work with the specification. Every version of the specification is now released with a change log to guide developers upgrading from a previous version, an Entity Relationship Diagram (ERD) to describe the whole data model visually, and a JSON Schema version of the model to be used in some automated development efforts. For groups who are using Java, Java Script, Python, R, or Drupal, there are community maintained libraries available that contain full BrAPI implementations ready to be added to some existing code. The BrAPI Test Server and the BRAVA validation tool are both still available to the community for testing purposes, and they have been maintained to support every version of the specification. Finally, the three new resources list pages on brapi.org advertise the other BrAPI compatible software available in the community. The BrAPPs list displays the 10 standalone, plug-and-play, applications available to the community. The servers list displays the 27 registered public data servers, their current status, and a form for registering additional servers. The compatible software list shows the 31 software applications that are BrAPI compliant, and again, a form for registering additional applications. The process for registering new items for these lists is completely voluntary, so the totals represent a lower bound for the number of BrAPPs, data repositories, and applications available in the the community.

Community Growth

The international BrAPI Community consists of the software developers, breeders, and related scientists working on BrAPI related projects and data sources. This community is what sustains the BrAPI project, building implementations, maintaining development tools, and providing input to enhance the specification. As the project has grown, so has the community. The BrAPI project started in June 2014 with less than ten people coming together to discuss the idea. Over the next nine years, the community has grown to between 200 and 250 members. The community mailing list has 208 members, and the BrAPI Slack workspace has 234 members. The project leadership uses the mailing list to broadcast newsletters, announcements, and updates to the community. The BrAPI Slack workspace allows members of the community to discuss specific topics and collaborate directly with each other.

The BrAPI Hackathons are a major staple of the BrAPI community. Twice a year, the community gathers to discuss the specification and collaborate on BrAPI related projects. This time is very valuable to the community; for some organizations, the hackathon is the only time during the year when they have time to work on anything related to BrAPI. During the COVID-19 pandemic, virtual hackathons took the place of in person events. While the virtual hackathons do not provide the same level of face-to-face time that is crucial to collaborative work, they did allow for more attendees to gather and share their opinions. The typical virtual hackathon has about twice as many registered attendees compared to an in-person hackathon. However, attendees have reported much more productive work time during the in-person events. As a compromise, going forward, the community leadership has decided to have one in-person hackathon and one virtual hackathon each year, to balance the advantages of both.

As the project has matured, a formal project leadership structure became increasingly important. As described in Figure 2, the project governance is divided into two groups. The Project Management team is responsible for the day to day operations of the project. The PI and Co-PI are responsible for the project funds, and responsible for hiring the BrAPI Project Coordinator who is paid by the project funds. The Advisory Board is a group of elected officials representing the community. The board is responsible for long term planning of the project, as well as quick decision making on behalf of the community. The two groups meet quarterly to report on progress and stay synchronized.

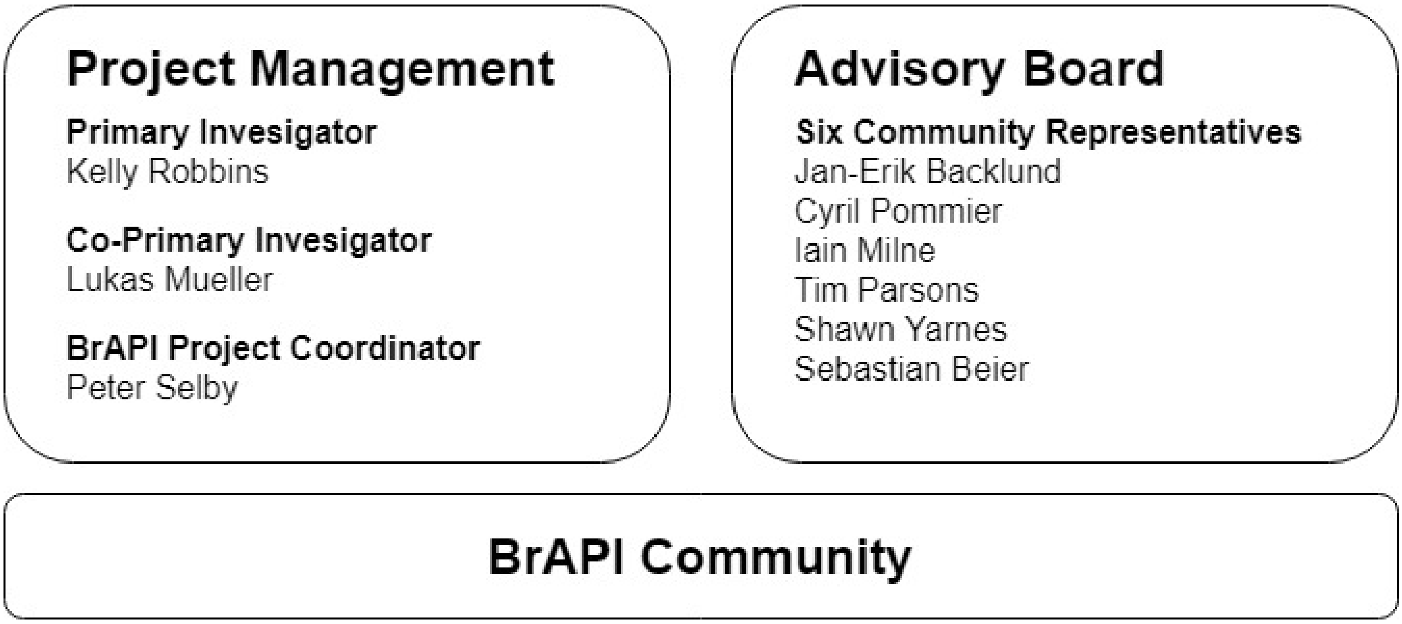

Figure 2: The formal governance structure of the BrAPI Project Community

Success Stories

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 5-10 years. These stories clearly illustrate all the different ways the BrAPI Standard can be used productively and in practice.

Data Collection

Field Book

Phenotypic data collection is an essential part of the breeding process. Historically, gathering data in the field was done with pen and paper, or perhaps some version of a digital spreadsheet. The abundance and prevalence of smart phones has allowed the Field Book mobile app to enhance data collection. Field Book can create well formed digital observation records from the moment they are taken. This can improve the efficiency of data collection and reduce human error.

BrAPI was able to take things a step further by automating the flow of data from the Field Book mobile app to a central database server. In this work flow, data is collected and stored completely digitally with little-to-no human involvement.

ClimMob

Not all data can be collected by a single person, or even by a single organization. ClimMob is a tool to easily allow citizen scientists to assist in the data collection process. Although this data may not be as detailed as a focused scientific program, it can be very useful to collect simple data from a wide range of locations and environments.

When it comes to BrAPI compatibility, ClimMob follows the same patterns established by Field Book. During a survey, all the farmer collected data is stored in a central ClimMob node. When the survey is complete, all the data is uploaded automatically via BrAPI to a central breeding database for long term storage and analysis.

ImageBreed

High-throughput phenotyping has been gaining significant traction lately as a way to collect lots of data very quickly. Image collection from unmanned arial and ground vehicles (UAVs and UGVs) are a great way to collect a lot of raw data all at once, then analyze it later. ImageBreed is a image collection pipeline tool to support regular use of UAVs and UGVs.

When the raw images have been processed through the standardization pipelines in ImageBreed, useful phenotypes can be extracted from the images. The BrAPI standard is used to push these phenotypes back to a central breeding database where they can be analyzed with other data. In addition to this, ImageBreed also has the option to use BrAPI to upload the raw images to the central breeding database, or any other BrAPI compatible long term storage service. The BrAPI models in the current version of the standard (V2.1) are rudimentary, but effective. The ImageBreed team has put in some work to enhance the BrAPI image data standards.

GridScore

Phenotypic data collection underpins scientific crop research and plant breeding. Knowledge gained from collected data and its analysis alongside data visualizations inform further phenotypic trials and ideally support research hypotheses. The importance of accuracy and efficiency in the collection of this data as well as the infrastructure to facilitate the flow of data from the field to a knowledge base cannot be underestimated. GridScore [2] is a modern mobile application for phenotypic observations that harnesses technological advancements in the area of mobile devices to enrich the data collection process.

BrAPI has further increased the value of GridScore by integrating it into the overarching workflow from trial creation, data collection, and its ultimate data storage for further processing. Specifically, trial designs as well as trait definitions can be imported into GridScore using BrAPI and a finalized trial can ultimately be exported via BrAPI to any compatible database.

Data Management

PHIS

The Hybrid Phenotyping Information System (PHIS[3]), based on the OpenSILEX framework, assumes responsibility for the systematic collection and management of data from phenotyping and high-throughput phenotyping experiments on a day-to-day basis. PHIS has the ability to efficiently store, organize and manage a wide range of data sets, including images, spectra and growth curves. This functionality extends to data at multiple spatial and temporal scales, from leaf to canopy, from a variety of sources such as field and greenhouse environments.

A key feature of PHIS is the unambiguous identification of all objects and traits within an experiment, establishing consistent relationships between them through the application of ontologies and semantics. This approach is designed to adapt to variations in experimental conditions, whether in the field or in controlled environments. PHIS's ontology-driven architecture emerges as a robust tool for integrating and managing data derived from diverse experiments and platforms, facilitating the creation of meaningful relationships between objects and augmenting datasets with relevant knowledge and metadata.

Furthermore, PHIS adheres to the Minimal Information About a Plant Phenotyping Experiment (MIAPPE[4]) and the Breeding API (BrAPI) standards.

The system recommends specific naming conventions, fostering a standardized approach for users to declare their resources. Notably, PHIS is widely adopted by various experimental platforms of the national PHENOME and European EMPHASIS infrastructure, serving as a hub for data management. Moreover, dedicated instances of PHIS have been established for the explicit purpose of resource sharing, encompassing projects, genetic resources, and variables, thereby fostering collaborative engagement and the dissemination of knowledge pertaining to studied concepts.

PHIS & BrAPI

PHIS offers a RESTful API designed to streamline interaction with data within a platform. Within this API, various services aligning with the Breeding API (BrAPI) standards have been implemented, encompassing the Core, Phenotyping, and Germplasm modules. Comprehensive documentation for these services is available on the PHIS Swagger interface. This integration with Swagger ensures that users can easily access, understand, and utilize the functionalities provided by the BrAPI-compliant web services, fostering transparency and facilitating effective engagement with the PHIS platform.

The design and development of PHIS have been meticulously tailored, incorporating the explicit constraint of aligning with BrAPI requirements. This intentional alignment ensures that PHIS adheres to the specified standards and protocols outlined by the Breeding API (BrAPI), thereby fostering seamless integration and compatibility with BrAPI-compliant systems and platforms. This prerequisite served as a substantial foundation for formalizing the data model, simultaneously facilitating compatibility with other standards, such as MIAPPE. Thus, by consciously incorporating BrAPI requirements into its structure, PHIS not only meets the phenotyping domain standards but also enhances its capacity for interoperability and effective collaboration within the broader context of plant breeding and related domains.

The fact that data within a PHIS instance can be queried through BrAPI services makes the indexing of PHIS in FAIDARE very easy to implement.

Indeed, as PHIS offers BrAPI-compliant Web Services, this greatly simplifies the integration and data exchange with other European information systems that handle phenotyping data. The adherence to BrAPI standards ensures a common interface and compatibility, facilitating seamless communication and collaboration between PHIS and other systems within the European context. This interoperability not only streamlines data sharing but also promotes a more cohesive and effective approach to managing and utilizing phenotyping data across diverse platforms and research initiatives in the European scientific community.

DeltaBreed

DeltaBreed is an open-source data management system designed and developed by Breeding Insight to support USDA-ARS specialty crop and animal breeders. DeltaBreed is a unified system for managing breeding data that connects a variety of BrAPI applications (see list below). BrAPI integration allows the complexity underlying interoperability to be hidden, shielding users from multifactorial differences between diverse 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.

DeltaBreed users need not be aware of BrAPI or the specifics of underlying applications 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 traits 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 will not be implemented until repeated observation handling workflows are established to differentiate and validate repeated observations, such as accidental repeats, overwrite requests, time-series observations, and repeated sub-entity measures. Users can expect DeltaBreed observation handling to become more seamless with future development.

DeltaBreed Connected Applications << Submission is expected April 2024. We may need to trim this aspirational list down to reality in final edits.>>

- BIMS https://www.breedwithbims.org/
- BrAPI Java Server https://test-server.brapi.org/brapi/v2/
- BrAPI Sync https://github.com/IntegratedBreedingPlatform/brapi-sync
- BreedBase https://breedbase.org/
- Diversity Arrays Technologies (DArT) genotyping services
- Field Book https://play.google.com/store/apps/details?id=com.fieldbook.tracker
- Gigwa https://gigwa.southgreen.fr/gigwa/
- Mr Bean https://github.com/AparicioJohan/MrBeanApp
- Pedigree Viewer https://github.com/solgenomics/BrAPI-Pedigree-Viewer

BMS

brapi sync and other things

Breedbase

Since the inception of BrAPI in 2014, the Breedbase group has been a heavy contributor to the BrAPI community.

BIMS

BIMS (Breeding Information Management System) [5] is a free, secure, and 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, data analysis and data archiving for their germplasm, phenotype, genotype, and image data. BIMS is currently implemented in five community databases, the Genome Database for Rosaceae [6], CottonGEN [7], the Citrus Genome Database, the Pulse Crop Database, and the Genome Database for Vaccinium, as well as a crop-independent website, https://breedwithbims.org. BIMS in these five community databases enables individual breeders to import publicly available data so that they can utilize public data in their breeding program. BIMS utilizes the Android App Field Book, enabling seamless data transfer between BIMS and the Field Book App through either files or BrAPI. Data transfer through BrAPI between BIMS and other resources such as BreedBase, GIGWA, and Breeder Genomics Hub is also on the way.

Germinate

Germinate [8] is an open-source plant genetic resources database that combines and integrates various kinds of plant breeding data including genotypic data, phenotypic trials data, passport data, images, geographic information and climate data into a single repository. Germinate is tightly linked to the BrAPI specification and supports a majority of BrAPI endpoints for querying, filtering and submission.

Germinate integrates and connects with other BrAPI-enabled tools such as GridScore for phenotypic data collection, Flapjack for genotypic data visualization and Helium for pedigree visualization, but, due to the nature of BrAPI, Germinate can act as a data repository for any BrAPI-compatible tool.

Thanks to the interoperability provided by BrAPI the need for manual data handling becomes a rarity with the direct benefit of faster data processing, fewer to no human errors, data security and integrity.

Federated Data Management Infrastructures

AGENT Portal

In the global system for ex situ conservation of plant genetic resources (PGR) [9], a total of ~5.8 million accessions are conserved in 1750 ex situ genebanks [10]. Unique and permanent identifiers in the form of DOIs are available for more than 1.7 million accessions [**doi:Food?** and Agriculture Organization (FAO) The Global Information System for PGRFA]. Each DOI is linked to some basic descriptive data that facilitates the use of these resources. Many DOIs are also linked to additional data from different domains or will be in the future. In order to answer questions on the global biological diversity of a plant species, on duplicate detection, on provenance tracking for the identification of genetic integrity, on the selection of the most suitable material for various purposes, including breeding and research, and to support further applications in data mining or AI, a data space beyond the most basic information is needed that includes genotypic and phenotypic data. In this context, the aim of the AGENT project (https://www.agent-project.eu/) funded by the European Commission is to develop a concept for the digital exploitation and activation of this GenRes data space via European ex situ genebanks according to the FAIR criteria [11] and to test it in practice using two important crops, barley and wheat. In two work packages, standards and technology for data interoperability will be developed to establish a genetic resources infrastructure, which regulates data acquisition of genotypic and phenotypic data, integrates and archives them and makes them accessible according to FAIR principles. To this end, 13 European genebanks and 5 bioinformatics centers are cooperating and have agreed on standards and protocols for (i) the data flow (see figure 3) and data formats [12] for central archiving of genotypic and phenotypic data.

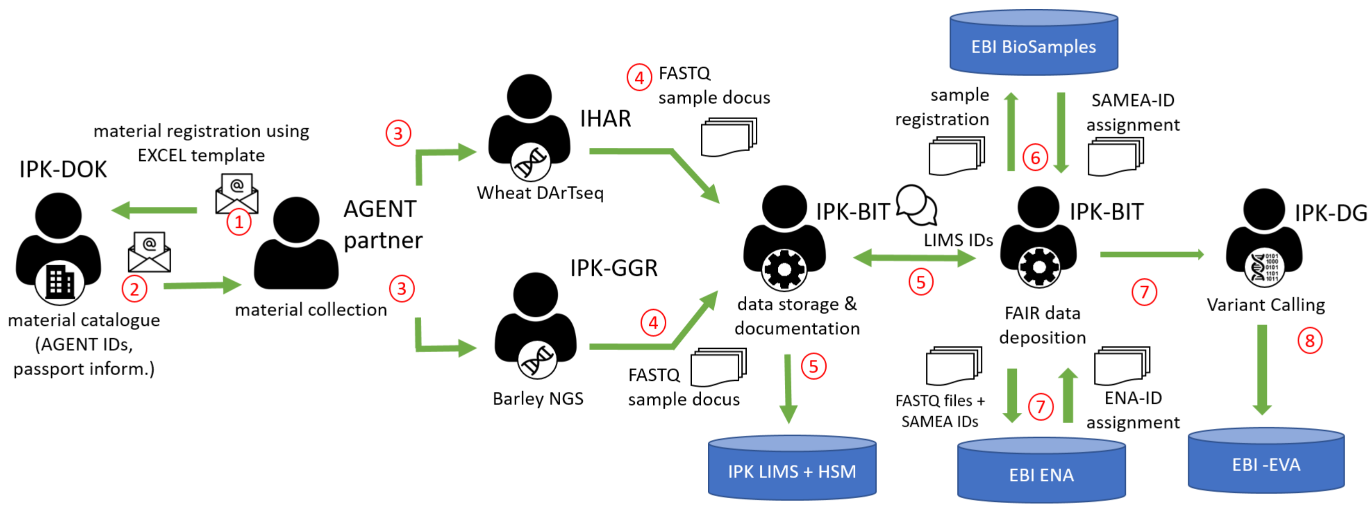

Figure 3: Figure Data flow of genotypic data from AGENT partner databases

The AGENT portal as described in more detail in section unlock the full potential of the biological material stored in genebanks around the globe by using FAIR international data standards and an open digital infrastructure for the management of plant genetic resources. The implemented BrAPI interface enables to mine current and historic genotypic and phenotypic information to drive the discovery of genes, traits and knowledge for future missions, complement existing information for wheat and barley and the new data standards and infrastructure to foster an improved management of PGR for other crop species across European genebanks.

For the joint research data infrastructure for the federation of collections of genotypic and phenotypic data from European gene banks and bioinformatics institutes, a AGENT portal (4) as database infrastructure for integrated plant genetic resources on ex-situ genebanks is being created. It provides, manual data exploration, machine-readable access via BrAPI and provide data to the cored data deposition resources at the European Bioinformatics Institute (EBI).

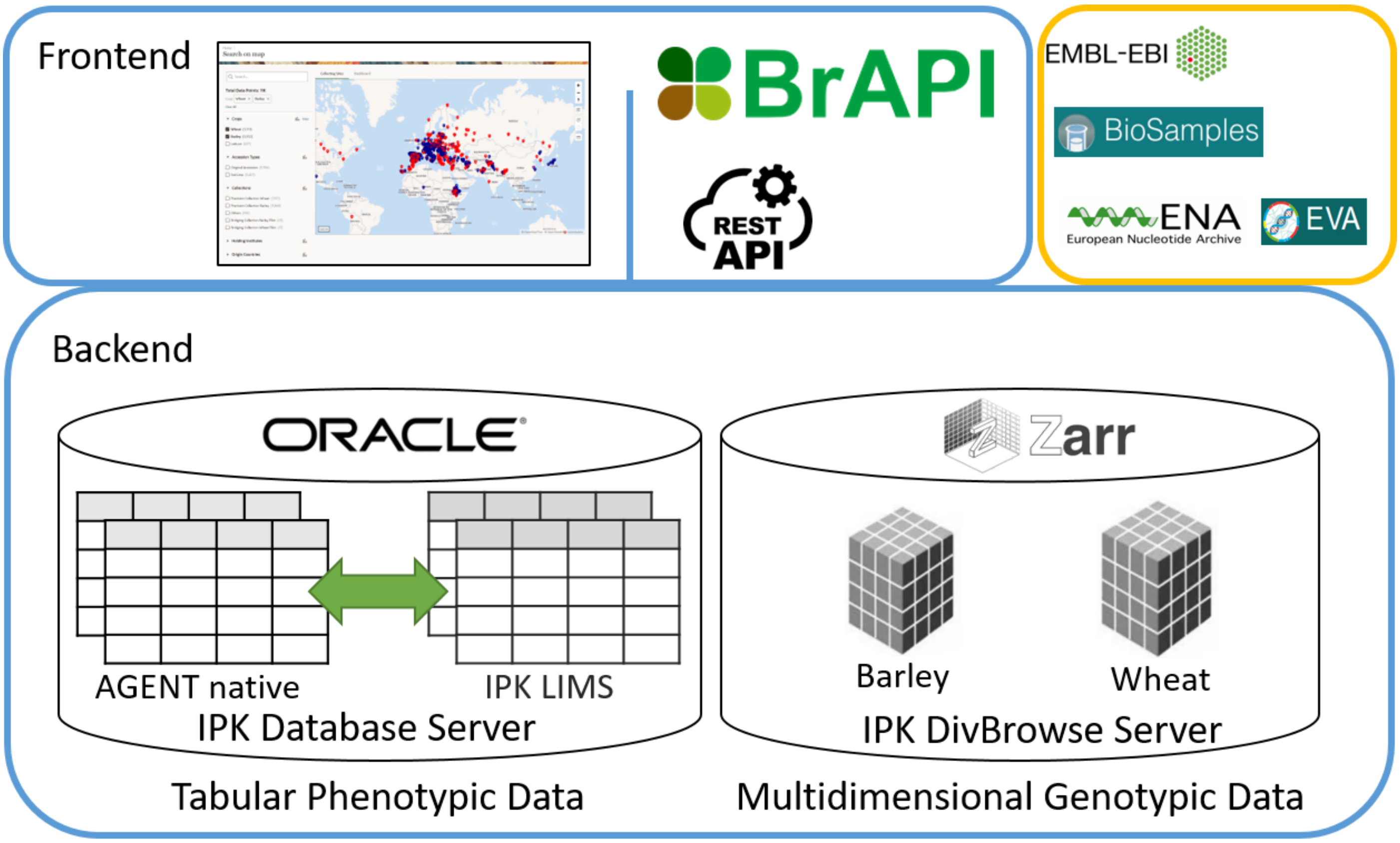

Figure 4: Figure AGENT Portal

The AGENT database backend aggregates curated and integrated passport data, phenotypic and genotypic data about wheat and barley accessions of 18 project partners are harmonized and integrated via BrAPI endpoints (https://github.com/AGENTproject/BrAPI) and explorable in a web portal (https://agent.ipk-gatersleben.de). The BrAPI endpoints were made available by scattered implementation. Genotyping data use DivBrowse [13] storage engine and BrAPI interface. Endpoints for sample data are implemented using AGENT database SQL to BrAPI broker service. To integrate those BrAPI endpoint provider into a single service and URL scheme, we work on their integration in a BrAPI proxy service. As next steps, we will expand BrAPI implementation to enable the integration of analysis pipelines in the AGENT portal, e.g. for genebank mining tools such as the FIGS+ pipeline developed by AGENT partner ICARDA [14]. Another perspective is to integrate the data collected in the AGENT project into the European Search Catalogue for Plant Genetic Resources (EURISCO) [15] and to implement BrAPI endpoints to make data on PGR collections in European genebanks programmatically accessible.

IPK-Genebank

Agrosystem Integration of germplasm collections in context of data trustee models among private economy and public research, integration of ex-situ genebanks (EU H2020 projects AGENT, INCREASING), integrated agrosystems and plant research infrastructure

MIAPPE “BrAPI to ISA” service

Since the release of BrAPI 1.3, efforts have been made to incorporate support for the Minimum Information About Plant Phenotyping Experiments (MIAPPE) standard into the specification [16]. This integration was finalized in BrAPI 2.0, resulting in full compatibility between the two standards. Consequently, BrAPI now encompasses all attributes necessary for MIAPPE compliance, adhering to standardized descriptions in accordance with MIAPPE guidelines. Leveraging BrAPI as a standardized RESTful web service API specification, we employ the ISA standard for storing metadata and phenotyping data in a standardized manner. This data is structured in the ISA-TAB file format and subjected to validation using the MIAPPE ISA configuration. The “BrAPI to ISA” service functions as a converter between BrAPI RESTful endpoints and ISA-TAB, facilitating the archiving of metadata and data and thereby enhancing data preservation and accessibility. The BrAPI2ISA tool is designed to be compatible with BrAPI 1.3, and we invite contributions from the community to extend support for the latest versions of BrAPI.

Data visualization

Flapjack

Flapjack [17] is a multi-platform desktop application for data visualization and breeding analysis (eg, pedigree verification, marker-assisted backcrossing and forward breeding) using high-throughput genotype data. Data can be easily 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, which 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 and work is underway to determine the best methods to exchange analysis results using future versions of the API.

Helium

Helium is a platform for reviewing and managing pedigree data. It is BrAPI enabled, so pedigree data can be viewed straight from a database.

Tassel

I don’t know much about Tassel or its BrAPI compliance. This is filler text for the layout of the manuscript.

Trait Selector BrAPP

BrAPPs are simple tools developed by the BrAPI community that are entirely reliant on BrAPI for their data requirements. This means a single BrAPI can be shared and used by many organizations, as long as those organizations have the standard BrAPI endpoints available. For example, the Trait Selector BrAPP is used to search and select useful traits, using a visual aid to help the user find exactly what they need. The BrAPP relies only on the BrAPI Trait endpoints to function, so any data resources with a the Trait endpoints available can add this tool to their system.

DArTView

DArTView is a desktop application for visualizing genotype variant data and looking for trends or correlations. It is newly BrAPI compatible and can use BrAPI as an input data source.

DivBrowse

DivBrowse is a BrAPI compatible tool for visualizing genotype variant data.

Analytics

QBMS

Modern breeding programs can utilize data management systems to maintain both phenotypic and genotypic data. Numerous systems are available for adoption. To fully leverage the benefits of digitalization in this ecosystem, breeders need to utilize 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.

To meet this demand, many breeding management systems have developed customized built-in analysis pipelines. However, these pipelines are static and may not accommodate evolving needs. As a result, API interfaces have been developed to facilitate data communication with other systems. The Breeding API (BrAPI) project specifies a standardized interface for plant phenotype/genotype databases, enabling them to share data with crop breeding applications. This promotes interoperability among plant breeding databases and allows third-party plugins to integrate with the ecosystem, delivering added value.

In the QBMS development team, we have identified a technical barrier between the breeding management systems' BrAPI interface and the scientists who create analysis scripts and pipelines. This barrier arises from the complexity of managing API backend processes, such as authentication, tokens, TCP/IP protocol, JSON format, pagination, stateless calls, asynchronous communication, database IDs, and more. To bridge this gap, we have developed the QBMS R package. This package abstracts the technical complexities, providing breeders (our end users) with stateful action verbs/functions that are familiar to them when navigating their GUI systems. This enables them to query and extract data into a standard data frame structure, consistent with their use of R language, one of the most common statistical tools in the breeding community.

Since its release on the official CRAN repository in October 2021, the QBMS R package has garnered over 7250 downloads. Several tools, such as MrBean, rely on the QBMS package as their source data adapter. Moreover, the community has started building extended solutions on top of it. QBMS can serve as a cornerstone in the breeding modernization revolution by providing access to actionable data and enabling the creation of dashboards to reduce the time between harvest and decision-making for the next breeding cycle.

Mr. Bean

Mr. Bean is an analytics and visualization tool. Mr. Bean has a data import tool which can pull data directly from a BrAPI compatible source.

G-Crunch

G-Crunch is an upcoming user-facing analysis tool that attempts to fill the space of simple, user driven analytics requests, with a generic user interface and the ability to swap out data sources and analysis tools. G-Crunch hopes to streamline repeatable, debuggable simple analytic requests and results. 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 (BrAPI-compliant) existing network of tools, which lowers the barrier to entry for adoption.

Samples and Genotypes

DArT Sample Submission

The DArT genotyping lab is heavily used world wide when it comes to plant genotyping. Developers at DArT have worked with the BrAPI community to establish a standard API for sending sample metadata to the lab before genotyping. This eliminates much of the human error involved with sending samples to en external lab.

MGIS

MGIS has germplasm and genotype data stored for many musa accessions. Through BrAPI, users are able to access this data directly from MusaBase, for use in specific experiments.

GIGWA

Gigwa is a JEE web application providing means to centralize, share, finely filter, and visualize high-throughput genotyping data [18]. Built on top of MongoDB, it is scalable and can support working smoothly with datasets containing billions of genotypes. Installable from docker images or all-in-one bundle archives, it is pretty 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/or a portal for exposing the data for genebanks and breeding programs for some CGIAR centers [19]. Thus, the amount of data hosted and made widely accessible using this system has kept growing over the last few years.

Gigwa developers have been involved in the BrAPI community since 2016 and took part in designing the genotype-related part of the API's specifications. Its first BrAPI-compliant features were designed for compatibility with the Flapjack visualization tool [17] and thus primarily turned it into a BrAPI datasource. Consequently, over time, Gigwa being the first and most reliable application implementing BrAPI-Genotyping server calls, local collaborators and even external partners used it as a reference solution to design a number of tools taking advantage of those features (e.g., BeegMac, SnpClust, QBMS). But further use-cases also required Gigwa to be able to consume data from other BrAPI servers, which led to also implement API-client features into the system. Thanks to all this work, a close collaboration was progressively established with the Integrated Breeding Platform team developing the widely used Breeding Management System, that ended up in both applications now being frequently deployed together, Gigwa pulling germplasm or sample metadata from BMS, and BMS displaying Gigwa-hosted genotypes within its own UI.

Client BrAPI libraries being available for R, community members typically write ad-hoc scripts syndicating data from multiple BrAPI sources (for instance phenotypes from a datasource and genotypes from another) in order to run various kinds of analyses such as GWAS, genomic selection or phylogenetic investigations. As a perspective, we may expect the most generic and widely-used of those pipelines to be at least publicly distributed, and possibly web-interfaced using solutions like R-Shiny in order to provide new, excitingly useful online services, based on Gigwa-hosted data.

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. Using a pangenome approach, each PHG stores haplotypes (the sequence of part of an individual chromosome) to represent the collected 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 access the crop databases either with direct calls to the PHG embedded server or indirectly using the rPHG library from an R environment. The PHG server accepts BrAPI endpoint queries to return information on sample lists and the variants used to define the graph's haplotypes. In addition, PHG users utilize the BrAPI variantsets endpoint query to return links to VCF files containing haplotype data. Work on the PHG is ongoing. We expect to support additional BrAPI endpoints that allow for slicing genotypic data based on samples and regions.

Data Portal

FAIDARE

FAIDARE (https://urgi.versailles.inrae.fr/faidare/) is a data discovery portal providing a biologist friendly search system over a global federation of 33 plant research databases. It allows to identify data resources using a full text approach completed with domain specific filters and to link back to the original database for visualization, analysis and download. For instance, it is possible to search for "wheat drought" then to refine the search to the "Triticum aestivum" taxon and yield component traits such as "Thousand Grain Weight". The indexed data types are very broad and include genomic features, such as genes or transposable elements, selected bibliography, QTL, markers, genetic variation studies, phenomic studies and plant genetic resources ie germplasm. This inclusiveness is achieved thanks to a two stage indexation data model. The most generic one provides basic search functionalities and relies on five fields : name, link back URL, data type, species and exhaustive description. The filtering is directly tied to some of those fields. Therefore, to provide more advanced filtering, FAIDARE is also providing a second stage indexation mechanism by taking advantage of BrAPi endpoints to get more detailed metadata on genotyping and phenotyping studies as well as germplasm. In parallel, FAIDARE provides a pre-visualization of germplasm and studies using dedicated cards. The indexation mechanism relies on a dedicated public software (https://github.com/elixir-europe/plant-brapi-etl-faidare) that allows data resources manager to request the indexation of there database using pull requests. It is able to extract data from any BrAPI 1.3 and 1.2 endpoint and development of BrAPI 2.x indexation will be initiated in 2025. Since not all databases are willing to implement BrAPI endpoints, we also provide the possibility to generate metadata as BrAPI json files, hence using the standard as a file exchange format. FAIDARE has been adopted by several communities and in particular in the ELIXIR and EMPHASIS european infrastructures. It is also used by the WheatIS of the Wheat-Initiative. Several databases are added each year to the FAIDARE global federation, allowing to increase both the portal and the BrAPI adoption.

Phenospex - HortControl

HortControl, developed by Phenospex, is a data repository. HortControl has a BrAPI implementation to be used to automate workflows and analytics software.

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 global entry point for 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) to PGRFA ex situ and in situ based on the Multi-Crop Passport Descriptors (MCPD), the 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 International Treaty and the Governing Body have repeatedly welcomed efforts on interoperability among germplasm information systems. In this context, the GLIS Portal adopted the Breeding API (BrAPI v1) in 2022. Integrating the 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.

FLORILÈGE (Gateway to French Plant Genetic Resources)

Designed primarily for the general public, Florilège provides access to all French plant biological resources centers. Its interface allows individuals to browse available plant accessions and gives them the possibility to order them. The listed accessions originate from 19 resources centers and concern around fifty plant species.

Florilège retrieves accession information from different BrAPI-compliant systems such as OLGA, an internal accessions management system, or FAIDARE. Leveraging the BrAPI implementation of these systems ensures standardized data retrieval from multiple sources, making the integration of new data sources that implement BrAPI an effortless process. The implementation of BrAPI is a prerequisite for the integration of any new database in Florilège.

Florilège is developed in Drupal 10, and uses xnttbrapi module (to easily connect to BrAPI compliant external databases).

BrAPI plug and play GraphQL based data-warehouse

Using the “Zendro” set of automatic software program-code generators (zendro-dev.github.io) a fully functional, efficient, and cloud-capable BrAPI data-warehouse has been created for the current version of the BrAPI data models. The resulting data-warehouse has two interfaces, one application programming interface implemented in the form of a GraphQL web-server and another intuitive point and click graphical user interface in the browser. Both provide secure access to data read and write functions for all BrAPI data models. These data administration methods comprise create, read, update, and delete (CRUD) functions that are standardized and accept the same parameters for all data models.

While data write access comprises both persisting single or multiple records, data read access is particularly rich in features and includes access to single records referred to by their id and access to multiple records selected by logical filters. In this, multiple records are paginated using the highly efficient cursor based pagination model as proposed in the GraphQL standard. Logical filters allow for exhaustive search queries, whose structure is highly intuitive and based around logical triplets in which a data model field is validated using an operator and a value, e.g. “Study name equals ‘xyz’”. In this 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 exactly for the data wanted.

Access security is implemented with the OAuth2 user authentication standard (datatracker.ietf.org/doc/html/rfc6749). Authorization is based on user roles and can be configured differently for each single data model read or write function.

The browser based graphical user interface is implemented in React.js with Next and exposes an intuitive and self explanatory set of functions for each data model. In the left a menu allows the user to access all BrAPI data models. Upon clicking on a model a table is shown which allows the user to paginate through all existing records, sort them by any column, search the records, add new records, or update or delete existing records, if the user role authorizes these functions. Record data can be inspected in a detail view and here relationships to other data records can be reviewed using the very same graphical visual representations. Breadcrumbs allow the user to navigate back and forth in the trail of relationships inspected. Finally, the generated graphical interface allows for the integration of interactive scientific plots and analysis tools written in JavaScript or WebAssembly.

The Zendro based BrAPI plug and play data-warehouse is capable of forming an efficient cloud of data servers. This is achieved simply by linking (URLs) other Zendro based warehouses that expose the same GraphQL API to the same data models, or a subset of data models. Any network of such Zendro GraphQL servers can be set up using this configuration approach. The code generated then exposes full access to all data records stored on any node of the network, while maintaining full security control at each node. Importantly, the warehouses are programmed in such a way that any number of data servers can be joined without loss of efficiency. Only the network connection speed and size of requested record sets influence the performance.

As explained, Zendro is a code generator and creates a fully functional data warehouse from input data model definitions, i.e. a schema. The schema is given in the form of special data model descriptions, in which each model is defined using JavaScript Object Notation (JSON). Each model is defined in its respective JSON file. A translator has been developed to create the Zendro schema from the BrAPI data model definitions. This ensures that Zendro can create plug and play data warehouses for future versions of the BrAPI with great ease, i.e. by translating the BrAPI models to Zendro input and subsequently running Zendro to create the plug and play warehouse.

Discussion

BrAPI for Breeders and Scientists

The BrAPI technical specification document is meant to be read and used by software developers. However, the purpose of the specification, and the community around it, is to make things faster, easier, and cheaper for the breeders and scientists working to make the world a better place. BrAPI offers a convenient path to automation and data integration for software tools in the breeding domain. All of the example use cases described above can be achieved with manual effort, moving and editing data files by hand. However, when the basic structure and flow of data becomes automated, breeders and scientists can spend less time on data management and more time focussing on the science, doing what they do best. For many, the ultimate goal is the development of a digital ecosystem: a collection of software tools and applications that can all work together seamlessly. In this digital ecosystem, data is collected digitally from the beginning, reducing as much human error as possible. The data is checked by quality control and stored automatically, then can be sent to any internal tool or external lab for further analysis with just the click of a button. This idea might sound too good to be true, but as more tools start sharing a universal data standard, automating data flow becomes easier, and the community gets closer to total interoperability.

Looking Ahead

The BrAPI specification will continue to grow, enabling more use cases and new types of data. These new use cases might include newer scientific techniques and technologies. Things like drone imaging data, spectroscopy, LIDAR, metabolomics, transcriptomics, high-throughput phenotyping, and machine learning analysis. All of these technologies can open new avenues for research and development of new crop varieties. All of these technologies also generate more data, and require data sharing between different software applications and data repositories. The BrAPI project leadership and community is committed to building the standards to support these new use cases as they arrive and become accepted by the scientific community. In fact, small groups within the BrAPI community have already start building generic data models and communication standards for many of the technologies listed above. These community efforts will eventually become part of the BrAPI standard in a future version of the specification document.

Conclusions and Impact

- High level summary of the project/consortium
- BrAPI is fitting into this gap, it doesn’t need to fit these other gaps
- Call to action - Join us!

References

1. **BrAPI—an application programming interface for plant breeding applications**
Peter Selby, Rafael Abbeloos, Jan Erik Backlund, Martin Basterrechea Salido, Guillaume Bauchet, Omar E Benites-Alfaro, Clay Birkett, Viana C Calaminos, Pierre Carceller, Guillaume Cornut, …
Bioinformatics (2019-03-23) https://doi.org/gjgxxr
DOI: 10.1093/bioinformatics/btz190 · PMID: 30903186 · PMCID: PMC6792114

2. **GridScore: a tool for accurate, cross-platform phenotypic data collection and visualization**
Sebastian Raubach, Miriam Schreiber, Paul D Shaw
BMC Bioinformatics (2022-06-06) https://doi.org/gtkcsv
DOI: 10.1186/s12859-022-04755-2 · PMID: 35668357 · PMCID: PMC9169276

3. **Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System**
Pascal Neveu, Anne Tireau, Nadine Hilgert, Vincent Nègre, Jonathan Mineau-Cesari, Nicolas Brichet, Romain Chapuis, Isabelle Sanchez, Cyril Pommier, Brigitte Charnomordic, … Llorenç Cabrera-Bosquet
New Phytologist (2019-01) https://doi.org/gm9b9j
DOI: https://doi.org/10.1111/nph.15385

4. **Enabling reusability of plant phenomic datasets with MIAPPE 1.1**
Evangelia A Papoutsoglou, Daniel Faria, Daniel Arend, Elizabeth Arnaud, Ioannis N Athanasiadis, Inês Chaves, Frederik Coppens, Guillaume Cornut, Bruno V Costa, Hanna Ćwiek-Kupczyńska, … Cyril Pommier
New Phytologist (2020-07) https://doi.org/gjqcmb
DOI: https://doi.org/10.1111/nph.16544

5. **The Breeding Information Management System (BIMS): an online resource for crop breeding**
Sook Jung, Taein Lee, Ksenija Gasic, BTodd Campbell, Jing Yu, Jodi Humann, Sushan Ru, Daniel Edge-Garza, Heidi Hough, Dorrie Main
Database (2021-01-01) https://doi.org/gtj95x
DOI: 10.1093/database/baab054 · PMID: 34415997 · PMCID: PMC8378516

6. **15 years of GDR: New data and functionality in the Genome Database for Rosaceae**
Sook Jung, Taein Lee, Chun-Huai Cheng, Katheryn Buble, Ping Zheng, Jing Yu, Jodi Humann, Stephen P Ficklin, Ksenija Gasic, Kristin Scott, … Dorrie Main
Nucleic Acids Research (2018-10-24) https://doi.org/gjqq8v
DOI: 10.1093/nar/gky1000 · PMID: 30357347 · PMCID: PMC6324069

7. **CottonGen: The Community Database for Cotton Genomics, Genetics, and Breeding Research**
Jing Yu, Sook Jung, Chun-Huai Cheng, Taein Lee, Ping Zheng, Katheryn Buble, James Crabb, Jodi Humann, Heidi Hough, Don Jones, … Dorrie Main
Plants (2021-12-18) https://doi.org/gqqwfm
DOI: 10.3390/plants10122805 · PMID: 34961276 · PMCID: PMC8705096

8. **From bits to bites: Advancement of the Germinate platform to support prebreeding informatics for crop wild relatives**
Sebastian Raubach, Benjamin Kilian, Kate Dreher, Ahmed Amri, Filippo M Bassi, Ousmane Boukar, Douglas Cook, Alan Cruickshank, Christian Fatokun, Noureddine El Haddad, …
Crop Science (2020-08-20) https://doi.org/gm66th

DOI: 10.1002/csc2.20248

9. **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**
Johannes MM Engels, Andreas W Ebert
Plants (2021-07-29) https://doi.org/gtq5n9
DOI: 10.3390/plants10081557 · PMID: 34451602 · PMCID: PMC8401695

10. **The Vulnerability of Plant Genetic Resources Conserved Ex Situ**
Yong-Bi Fu
Crop Science (2017-07-27) https://doi.org/gbzbnd
DOI: 10.2135/cropsci2017.01.0014

11. **The FAIR Guiding Principles for scientific data management and stewardship**
Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, … Barend Mons
Scientific Data (2016-03-15) https://doi.org/bdd4
DOI: 10.1038/sdata.2016.18 · PMID: 26978244 · PMCID: PMC4792175

12. **Recommendations for the formatting of Variant Call Format (VCF) files to make plant genotyping data FAIR**
Sebastian Beier, Anne Fiebig, Cyril Pommier, Isuru Liyanage, Matthias Lange, Paul J Kersey, Stephan Weise, Richard Finkers, Baron Koylass, Timothee Cezard, … Uwe Scholz
F1000Research (2022-05-19) https://doi.org/gtq5n7
DOI: 10.12688/f1000research.109080.2 · PMID: 35811804 · PMCID: PMC9218589

13. **DivBrowse—interactive visualization and exploratory data analysis of variant call matrices**
Patrick König, Sebastian Beier, Martin Mascher, Nils Stein, Matthias Lange, Uwe Scholz
GigaScience (2022-12-28) https://doi.org/gtq5n6
DOI: 10.1093/gigascience/giad025 · PMID: 37083938 · PMCID: PMC10120423

14. **Genebank mining with FIGS, the Focused Identification of Germplasm Strategy**
Ken Street, Ken Street
Unknown (2017) https://doi.org/gtq5n8
DOI: 10.22004/ag.econ.266624

15. **EURISCO update 2023: the European Search Catalogue for Plant Genetic Resources, a pillar for documentation of genebank material**
Pragna Kotni, Theo van Hintum, Lorenzo Maggioni, Markus Oppermann, Stephan Weise
Nucleic Acids Research (2022-10-03) https://doi.org/gqxshq
DOI: 10.1093/nar/gkac852 · PMID: 36189883 · PMCID: PMC9825528

16. **Enabling reusability of plant phenomic datasets with MIAPPE 1.1**
Evangelia A Papoutsoglou, Daniel Faria, Daniel Arend, Elizabeth Arnaud, Ioannis N Athanasiadis, Inês Chaves, Frederik Coppens, Guillaume Cornut, Bruno V Costa, Hanna Ćwiek-Kupczyńska, … Cyril Pommier
New Phytologist (2020-04-25) https://doi.org/gjqcmb
DOI: 10.1111/nph.16544 · PMID: 32171029 · PMCID: PMC7317793

17. **Flapjack—graphical genotype visualization**
Iain Milne, Paul Shaw, Gordon Stephen, Micha Bayer, Linda Cardle, William TB Thomas, Andrew J Flavell, David Marshall

Bioinformatics (2010-10-18) https://doi.org/cdvnt8
DOI: 10.1093/bioinformatics/btq580 · PMID: 20956241 · PMCID: PMC2995120

18. **Gigwa v2—Extended and improved genotype investigator**
Guilhem Sempéré, Adrien Pételc, Mathieu Rouard, Julien Frouin, Yann Hueber, Fabien De Bellis, Pierre Larmande
GigaScience (2019-05-01) https://doi.org/gtp5bz
DOI: 10.1093/gigascience/giz051 · PMID: 31077313 · PMCID: PMC6511067

19. **A digital catalog of high-density markers for banana germplasm collections**
Mathieu Rouard, Julie Sardos, Guilhem Sempéré, Catherine Breton, Valentin Guignon, Ines Van den Houwe, Sebastien C Carpentier, Nicolas Roux
PLANTS, PEOPLE, PLANET (2021-03) https://doi.org/gtp5bx
DOI: 10.1002/ppp3.10187