AGENT

Global Access to Plant Genetic Resources

Introduction to AGENT API

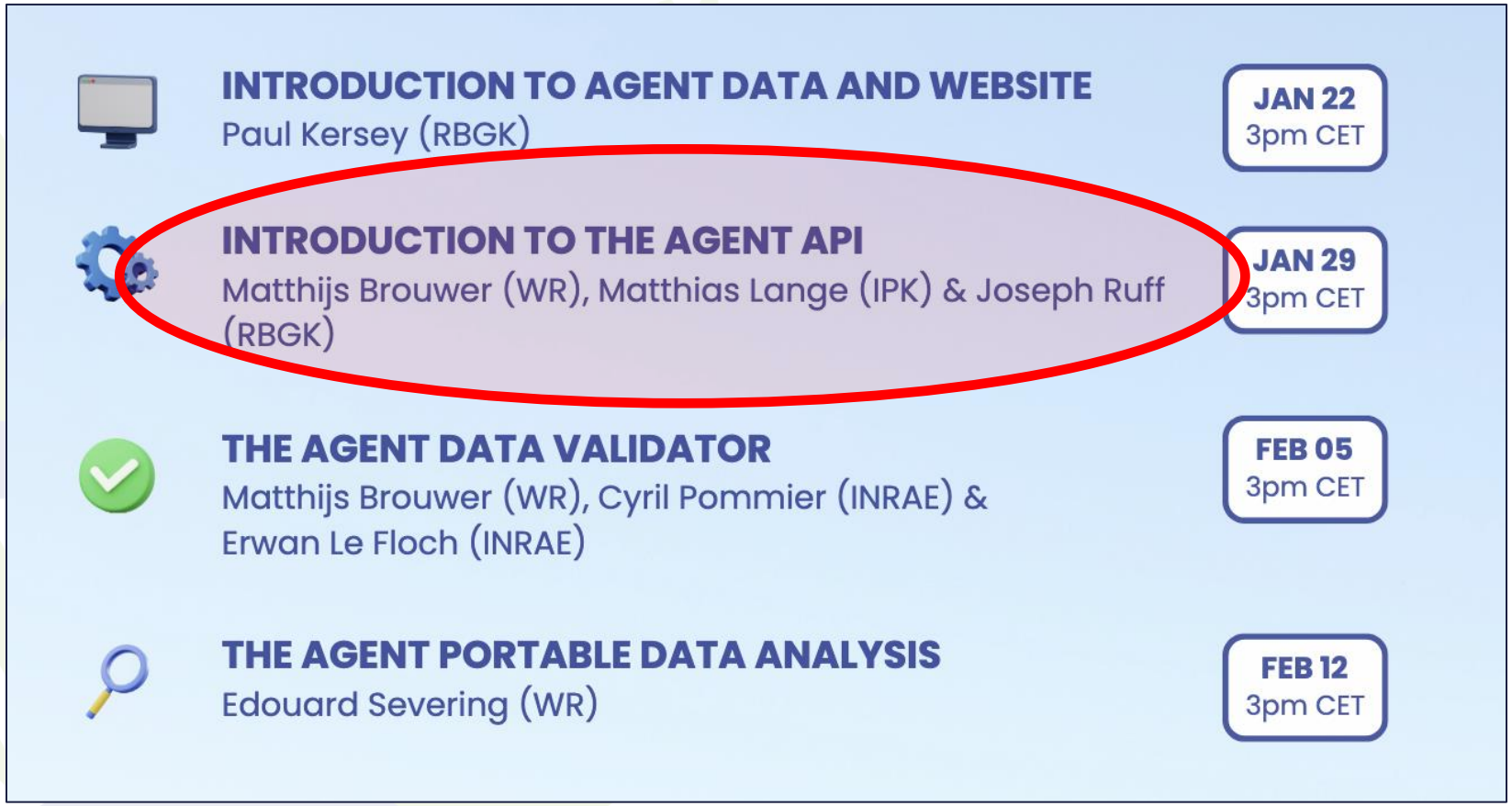
- Matthijs Brouwer
- Wageningen University & Research (WUR)

- 29 January 2025

WAGENINGEN
UNIVERSITY & RESEARCH

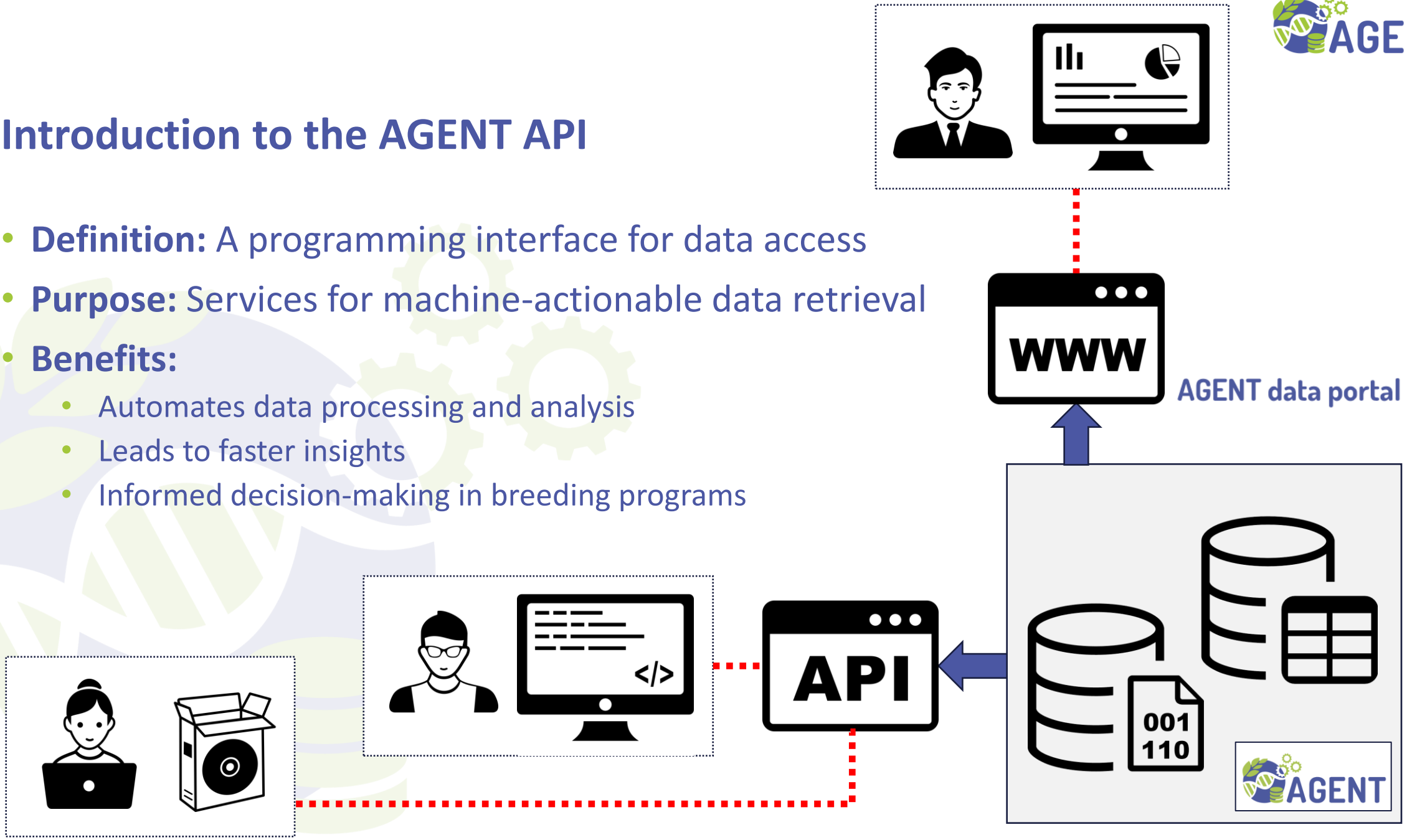
The AGENT project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 862613.

AGENT Webinar Series



Introduction to the AGENT API

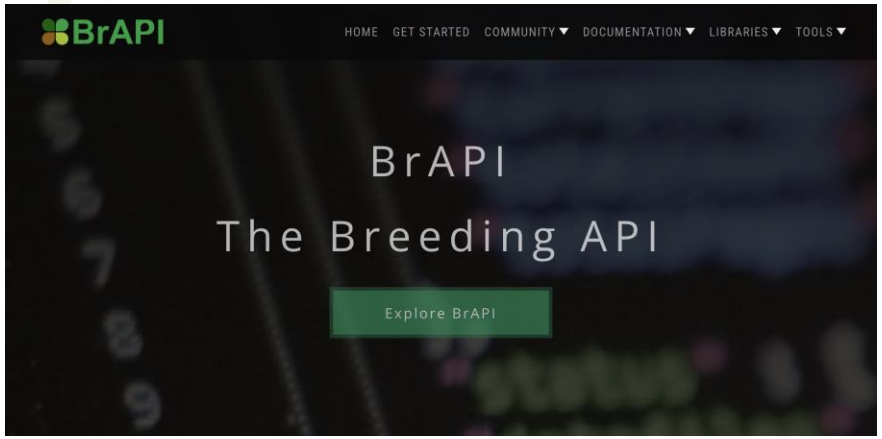
- **Definition:** A programming interface for data access
- **Purpose:** Services for machine-actionable data retrieval
- **Benefits:**
  - Automates data processing and analysis
  - Leads to faster insights
  - Informed decision-making in breeding programs



Adoption of Breeding API (BrAPI)

- **Goal:** Ensure a sustainable and internationally accepted API
- **Alignment:** With phenotyping and genotyping standards
- **Context:** Managing plant genetic resources in a big data environment

See https://brapi.org/



What is BrAPI?

- 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.

Advisory Board

Cyril Pommier

Advisory Board Member

INRAE URGI

Dr Stephan Weise

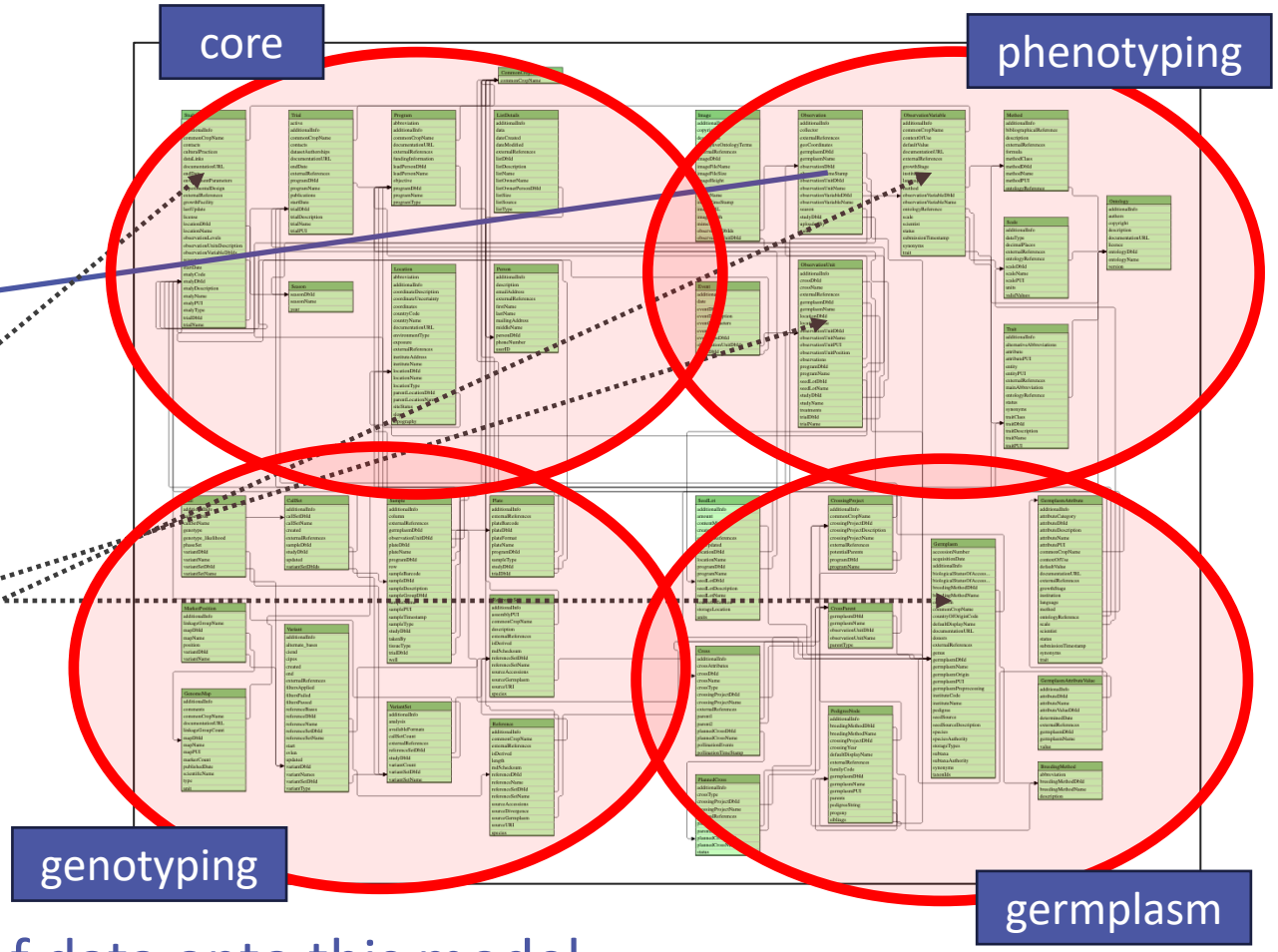
Head of Genebank Documentation / EURISCO Coordinator

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

What is BrAPI?

Model related entities

Observation
additionalInfo
collector
externalReferences
geoCoordinates
germplasmDbId
germplasmName
observationDbId
observationTimeStamp
observationUnitDbId
observationUnitName
observationVariableDbId
observationVariableName
season
studyDbId
uploadedBy
value



Always mapping of data onto this model

BrAPI is compatible with several community standards including MCPD, MIAPPE, GA4GH Variants Schema, GeoJSON, and the Crop Ontology

What is BrAPI?

- RESTful web service
- BrAPI endpoints
- Get the entities from the model

```
Observation:
  properties:
    additionalInfo:
      additionalProperties:
        type: string
      description: A free space containing any
        additional information related to a particular
        object. A data source may provide any JSON
        object, unrestricted by the BrAPI specification.
      type: object
    collector:
      description: The name or identifier of the entity
        which collected the observation
      example: 917d3ae0
      type: string
    externalReferences:
      description: An array of external reference ids.
        These are references to this piece of data in an
        external system. Could be a simple string or a
```

https://www.plantbreeding.wur.nl/agent/brapi/observations

JSON Raw Data Headers

Save Copy Pretty Print


```
{
  "@context": [
    "https://brapi.org/jsonld/context/metadata.jsonld"
  ],
  "metadata": {
    "datafiles": null,
    "status": [],
    "pagination": {
      "currentPage": 0,
      "pageSize": 1000,
      "totalCount": 994166,
      "totalPages": 995
    }
  },
  "result": {
    "data": [
      {
        "additionalInfo": null,
        "germplasmDbId": 121390,
        "observationDbId": 157737,
        "observationTimeStamp": null,
        "observationUnitDbId": 157737,
        "observationVariableDbId": 10121,
        "observationVariableName": "Days To Heading",
        "studyDbId": 8383,
        "uploadedBy": "KUMAR@IPK-GATERSLEBEN.DE",
        "value": "220"
      },
      {
        "additionalInfo": null,
        "germplasmDbId": 110857,
        "observationDbId": 157738,
```


pagination

BrAPI

JSON format

What is BrAPI?

"programming interface for data access"

```
import requests
url = "https://www.plantbreeding.wur.nl/agent/brapi/observations"
response = requests.get(url)
data = response.json()
```

Metadata - pagination

```
data["metadata"]["pagination"]
```

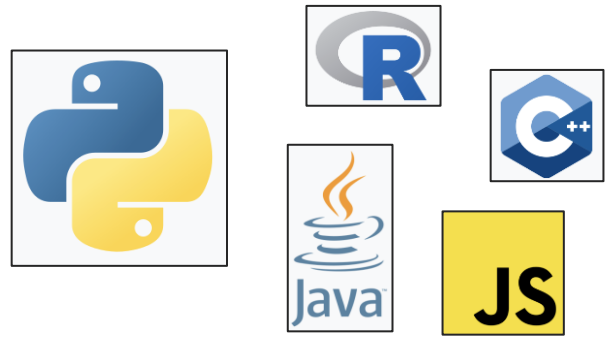
```
{'currentPage': 0, 'pageSize': 1000, 'totalCount': 994166, 'totalPages': 995}
```

Data

```
observations = pd.DataFrame(data["result"]["data"])
```

```
import requests
url = "https://www.plantbreeding.wur.nl/agent/brapi/observations"
headers = {"Authorization": "Bearer wi4u6w2dm361570l9n9ftq4tqvuq64yq"}
response = requests.get(url,headers=headers)
data = response.json()
```

HTTP error code 403
Authorization token required



Collect data from all pages with additional request

What is BrAPI?

"programming interface for data access"

```
import requests
url = "https://www.plantbreeding.wur.nl/agent/brapi/observations"
response = requests.get(url)
data = response.json()
```

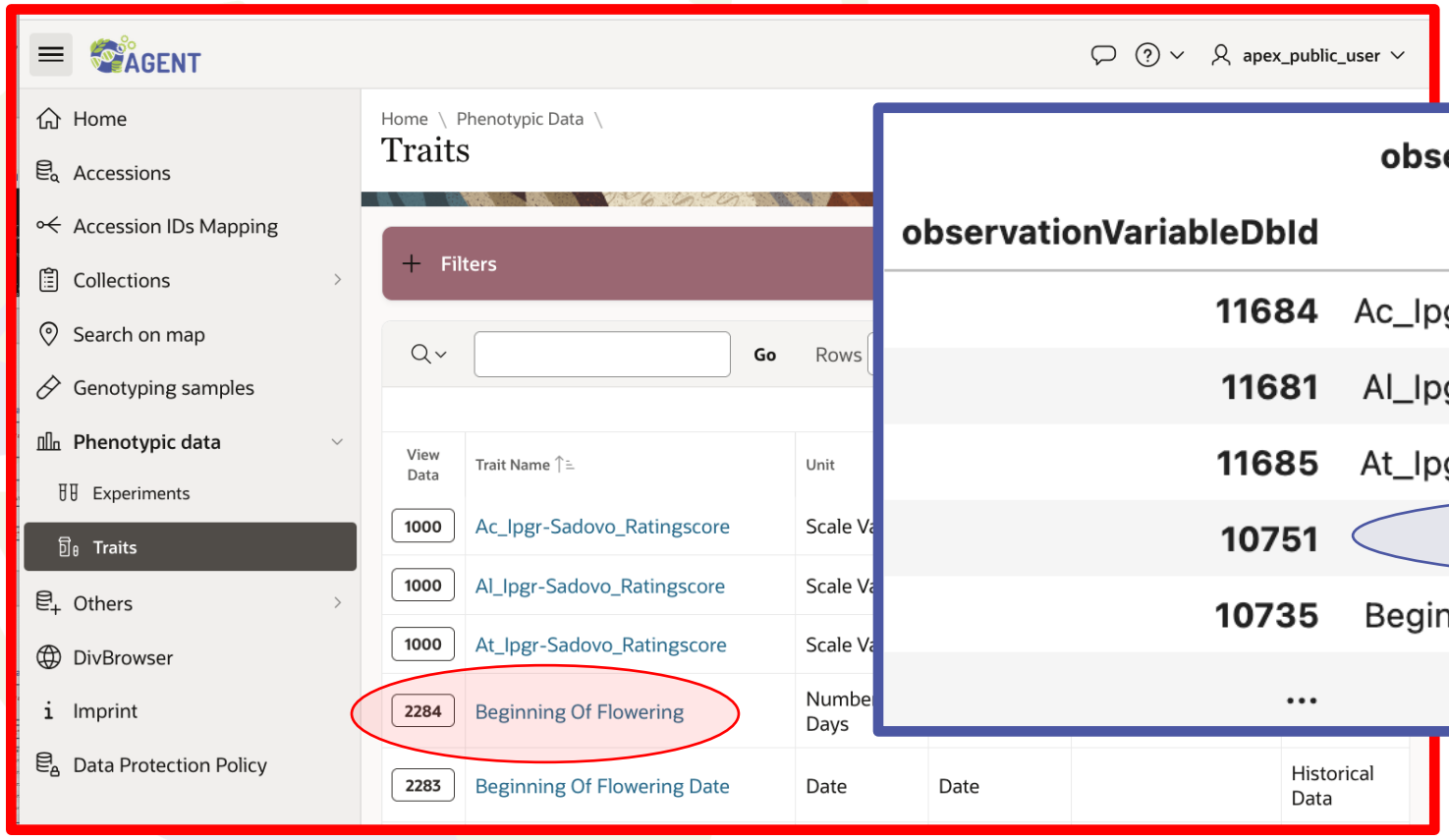
```
observations = pd.DataFrame(data["result"]["data"])
```

germplasmDbId	observationDbId	observationTimeStamp	observationUnitDbId	observationVariableDbId	observationVariableName	studyDbId
AB_39487	157737	None	157737	10121	Days To Heading	8383
AB_56176	157738	None	157738	10121	Days To Heading	8383
AB_49454	157739	None	157739	10121	Days To Heading	8383
AB_42868	157740	None	157740	10121	Days To Heading	8383
AB_46204	157741	None	157741	10121	Days To Heading	8383
...	...	...	...	...	...	...

What is BrAPI?

AGENT data portal

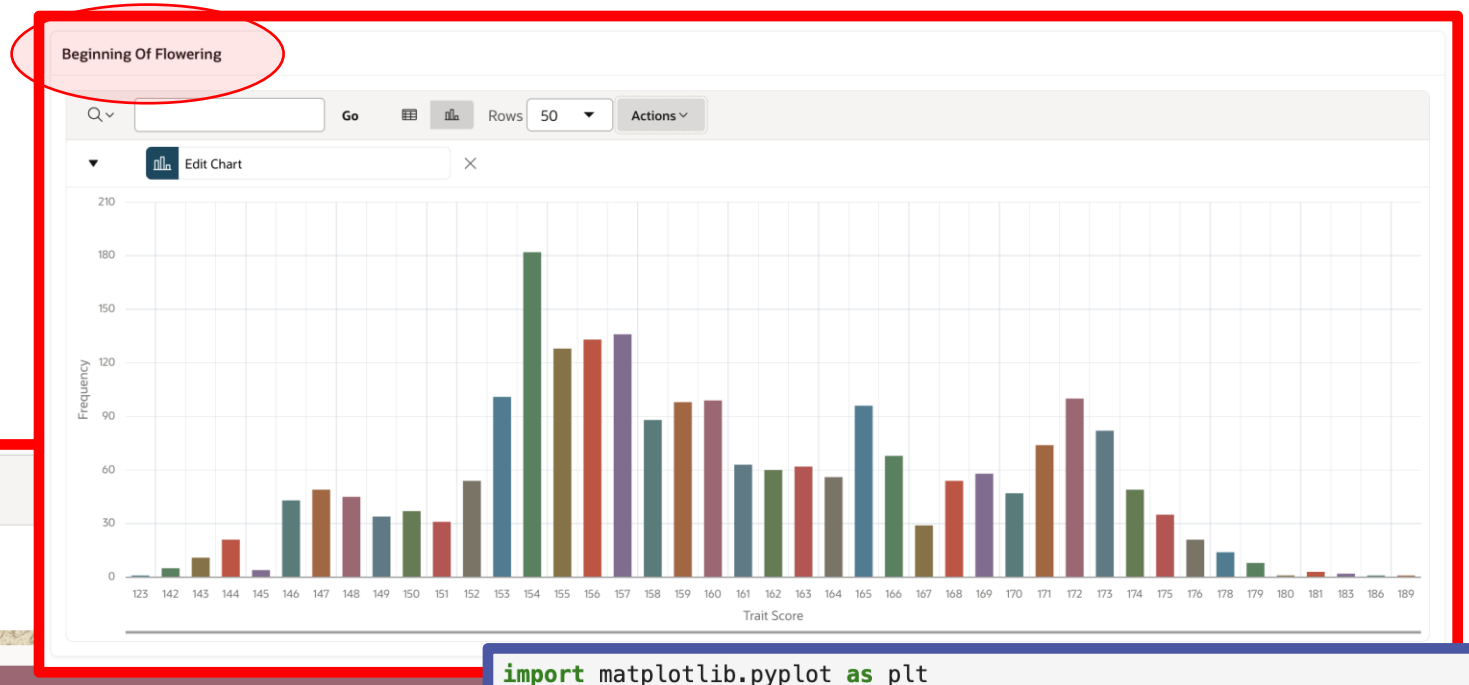
```
observations.groupby(["observationVariableDbId"], as_index=False).agg(
    {"observationVariableDbId": "first",
     "observationVariableName": "first",
     "observationDbId": "count"}
).set_index("observationVariableDbId"
).sort_values("observationVariableDbId")
```

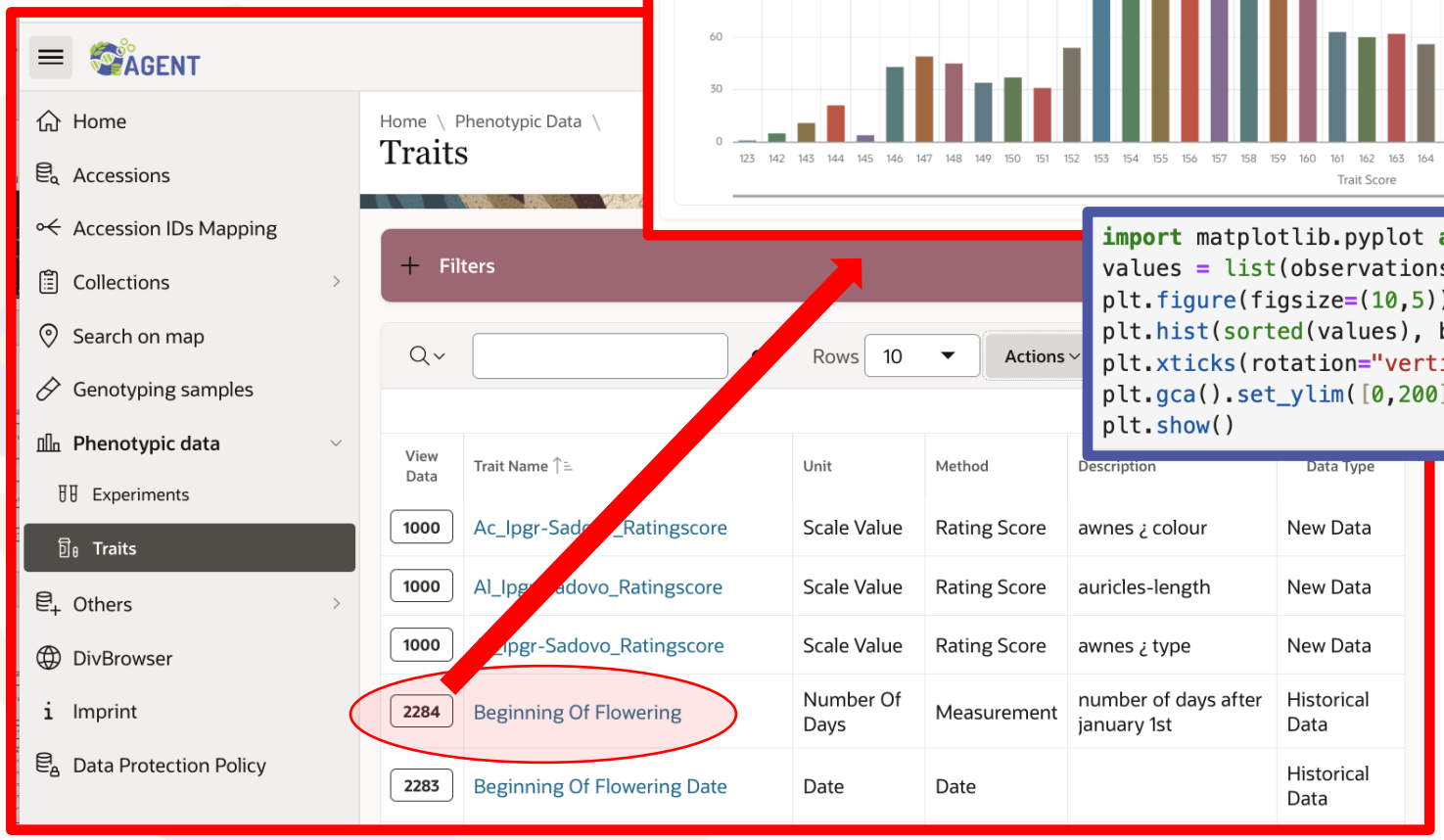


observationVariableDbId	observationVariableName	observationDbId
11684	Ac_Ipgr-Sadovo_Ratingscore	1000
11681	Al_Ipgr-Sadovo_Ratingscore	1000
11685	At_Ipgr-Sadovo_Ratingscore	1000
10751	Beginning Of Flowering	2284
10735	Beginning Of Flowering Date	2283
...	...	...

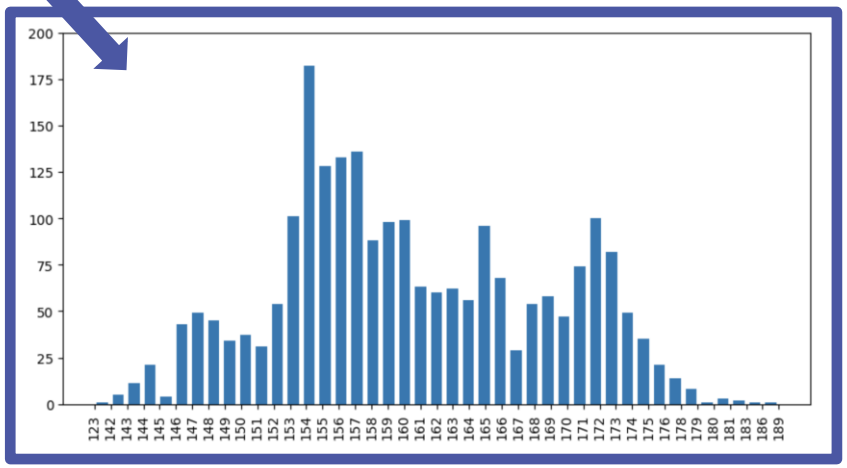
What is BrAPI?

AGENT data portal





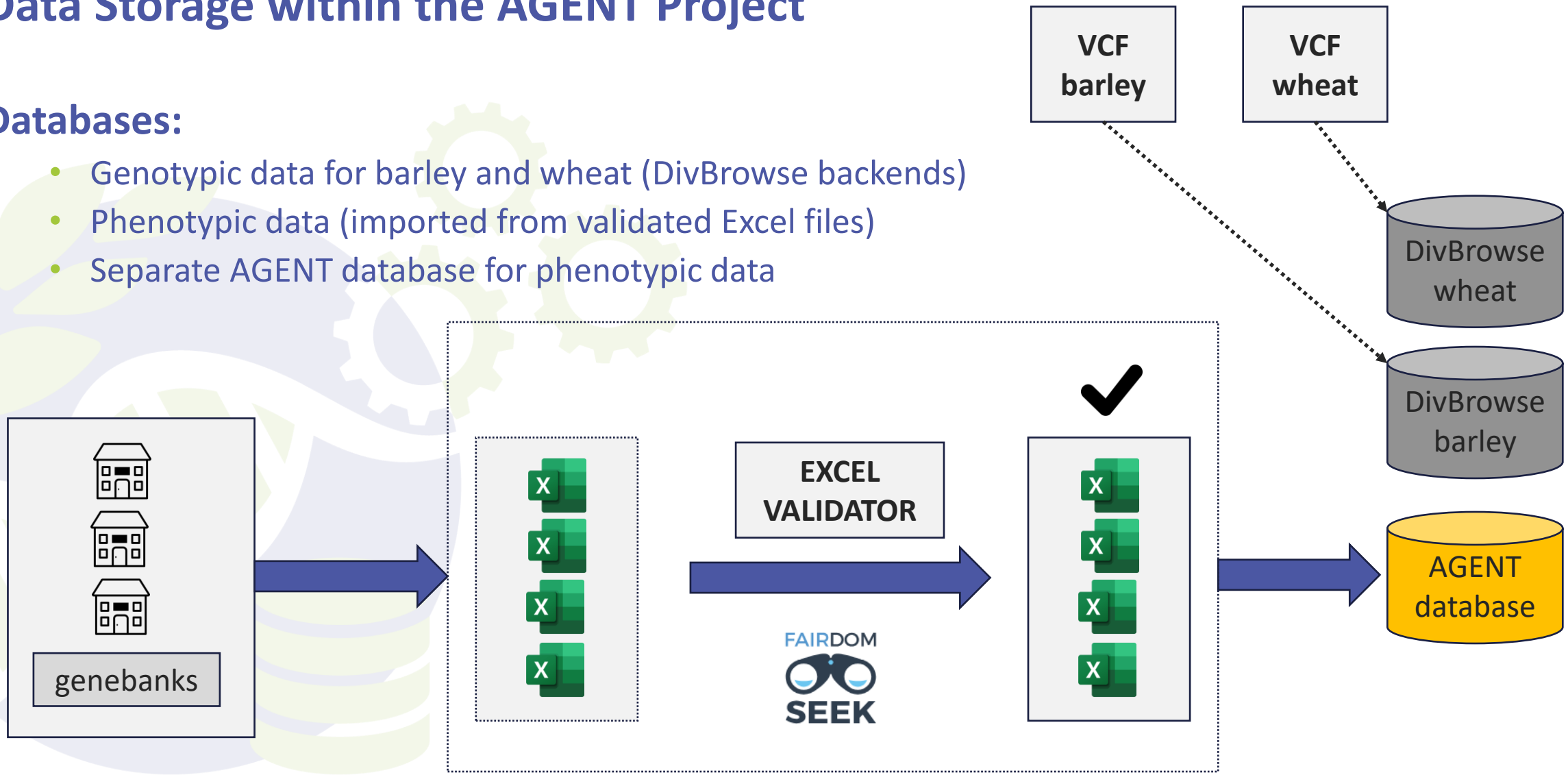

```
import matplotlib.pyplot as plt
values = list(observations[observations["observationVariableDbId"]==10751]["value"])
plt.figure(figsize=(10,5))
plt.hist(sorted(values), bins=43, rwidth=0.7)
plt.xticks(rotation="vertical")
plt.gca().set_ylim([0,200])
plt.show()
```



Data Storage within the AGENT Project

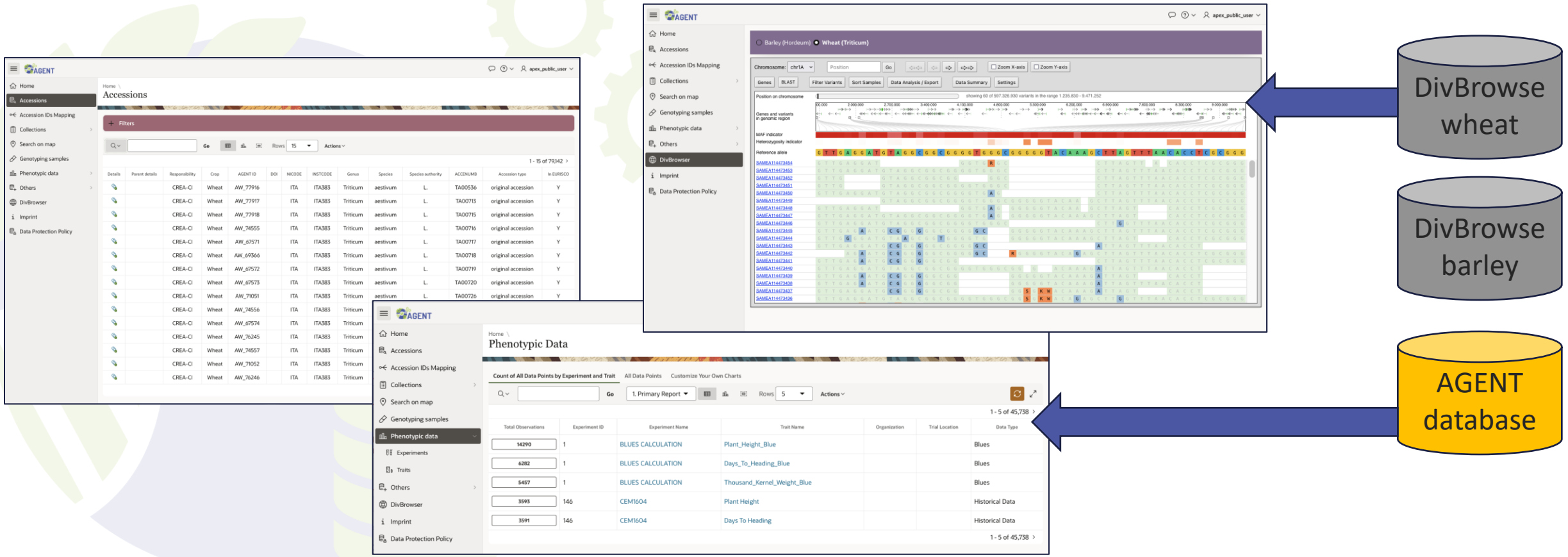
Databases:

- Genotypic data for barley and wheat (DivBrowse backends)
- Phenotypic data (imported from validated Excel files)
- Separate AGENT database for phenotypic data



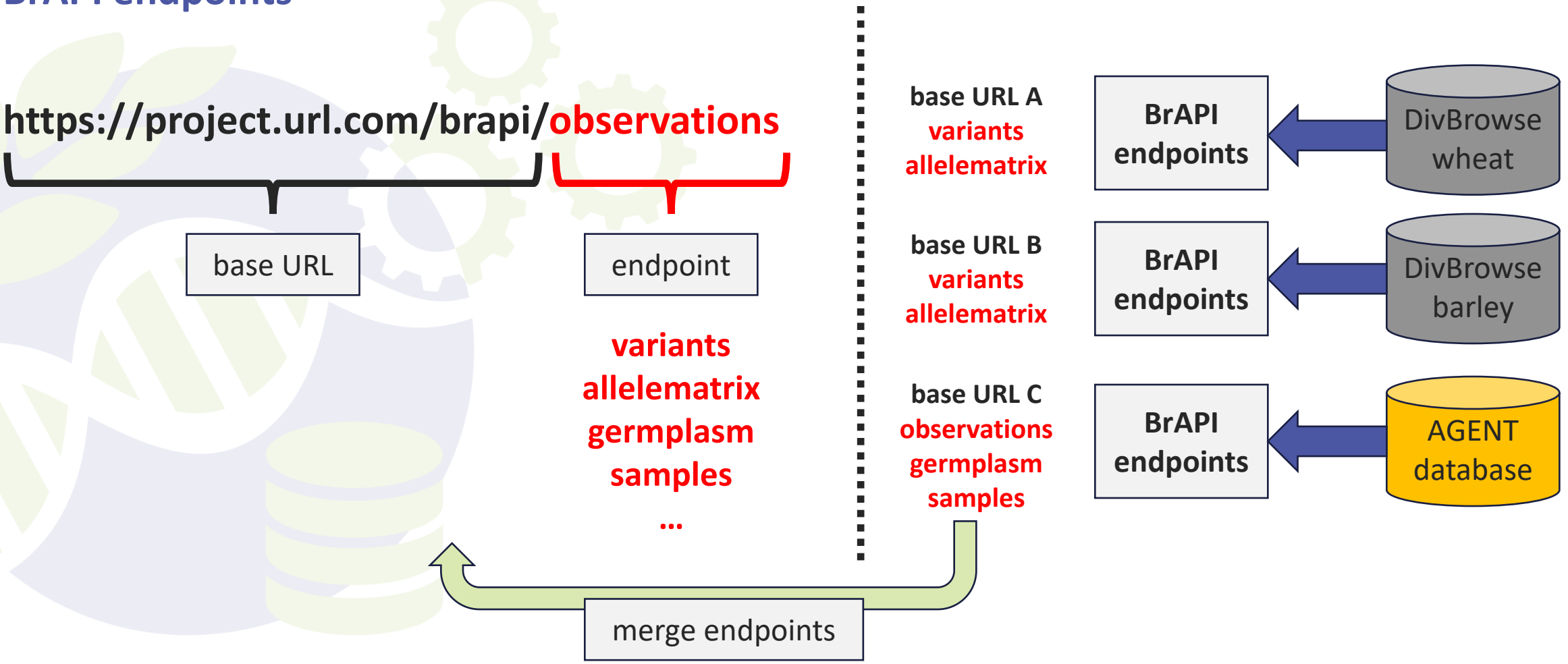
Data Storage within the AGENT Project

User Access: Through the AGENT portal



Data Storage within the AGENT Project

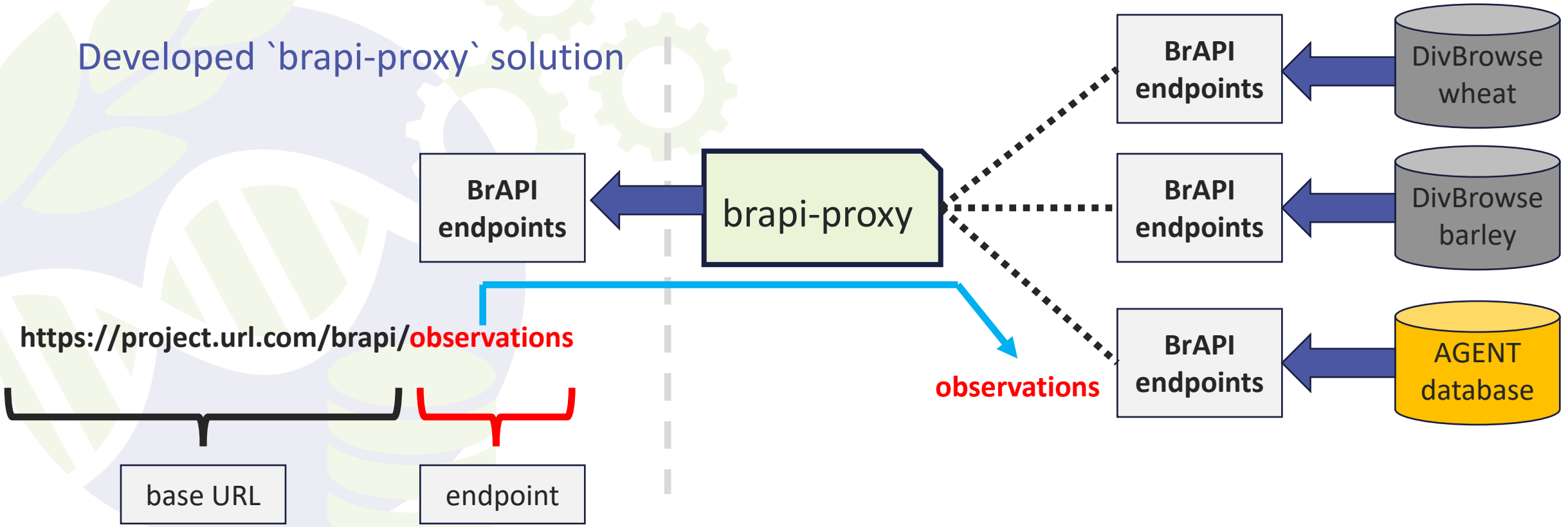
BrAPI endpoints



BrAPI Endpoints

Integration Challenge: merge BrAPI endpoints

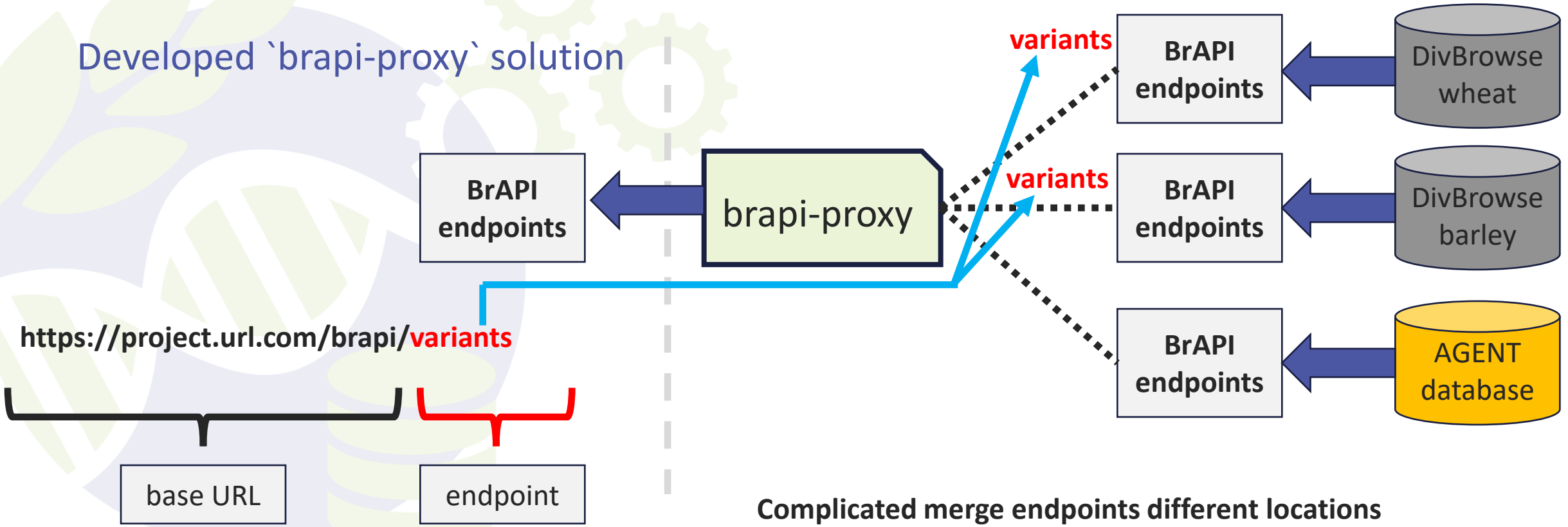
Developed `brapi-proxy` solution



BrAPI Endpoints

Integration Challenge: merge BrAPI endpoints

Developed `brapi-proxy` solution



Complicated merge endpoints different locations

BrAPI Endpoints

base URL | endpoint | parameters

Example: **https://project.url.com/brapi/variants?page=5&pageSize=100**

variants?page=5&pageSize=100


```
{
  "metadata": {
    "datafiles": [],
    "pagination": {
      "currentPage": 5,
      "pageSize": 100,
      "totalCount": 216956,
      "totalPages": 2170
    }
  },
  "result": {
    "data": [
      {
        "variantDbId": "chr1H:4521471",
        "additionalInfo": {},
        "referenceBases": "C",
        "alternateBases": [
          "T"
        ],
        "ciend": [],
```


DivBrowse barley

variants?page=0&pageSize=1


```
{
  "metadata": {
    "datafiles": [],
    "pagination": {
      "currentPage": 0,
      "pageSize": 1,
      "totalCount": 151488,
      "totalPages": 151488
    }
  },
  "result": {
    "data": [
      {
        "variantDbId": "Chr1A:1235830",
        "additionalInfo": {},
        "referenceBases": "G",
        "alternateBases": [
          "T"
        ],
        "ciend": [],
```


DivBrowse wheat

prefix


```
{
  "metadata": {
    "datafiles": null,
    "status": [],
    "pagination": {
      "currentPage": 5,
      "pageSize": 100,
      "totalCount": 368444,
      "totalPages": 3685
    }
  },
  "result": {
    "data": [
      {
        "variantDbId": "barley:chr1H:4521471",
        "additionalInfo": {},
        "referenceBases": "C",
        "alternateBases": [
          "T"
        ],
        "ciend": [],
```

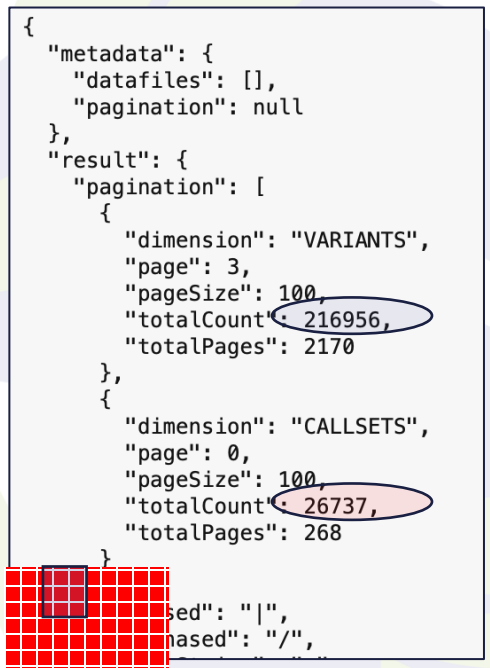

brapi-proxy

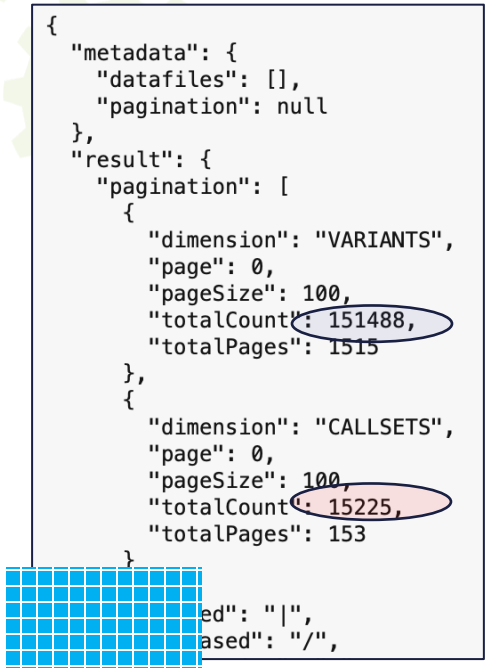
BrAPI Endpoints

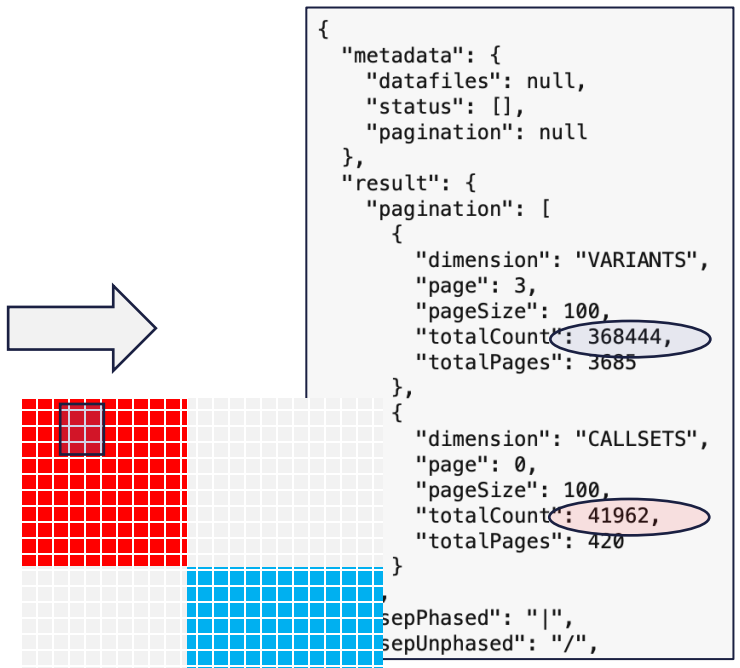
base URL | endpoint | parameters

Example: **https://project.url.com/brapi/allelematrix?dimensionVariantPage=3**

allelematrix?dimensionVariantPage=3 **allelematrix?dimensionVariantPage=0**







DivBrowse barley

DivBrowse wheat

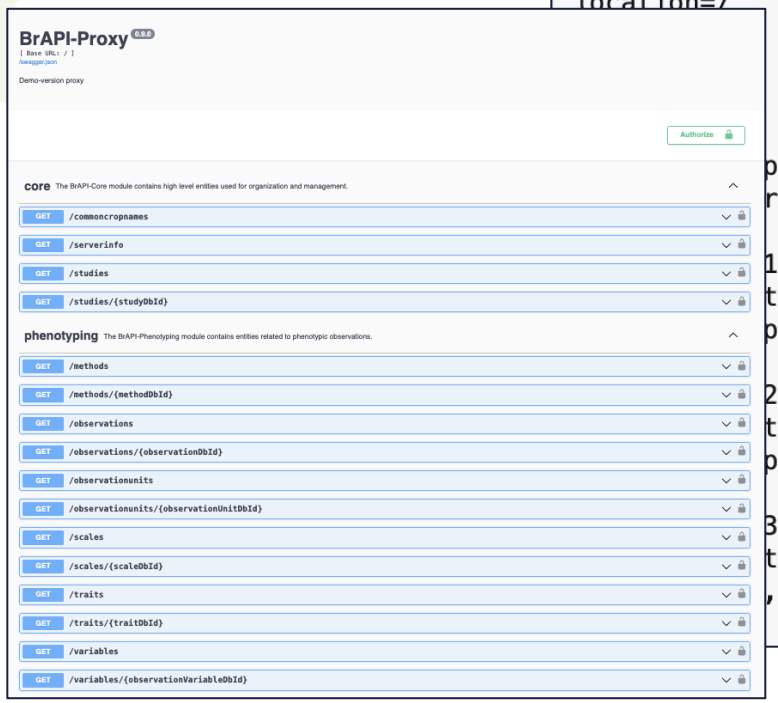
brapi-proxy

General BrAPI Proxy Solution

- **Purpose:** Integrate scattered endpoints into a single location
- **Functionality:** Combines and merges existing endpoints
- **Features:**
  - prefix identifiers
  - usage authorization token

https://pypi.org/project/brapi-proxy/

```
pip install brapi_proxy
```




```
[brapi]
port=8080
host=0.0.0.0
location=/
option=Demo-version proxy
rAPI-Proxy

1]
test-server.brapi.org/brapi/v2
ping.*

2]
test-server.brapi.org/brapi/v2
ping.*

3]
test-server.brapi.org/brapi/v2
,phenotyping.*,germplasm.*
```

General BrAPI Proxy Solution

Features

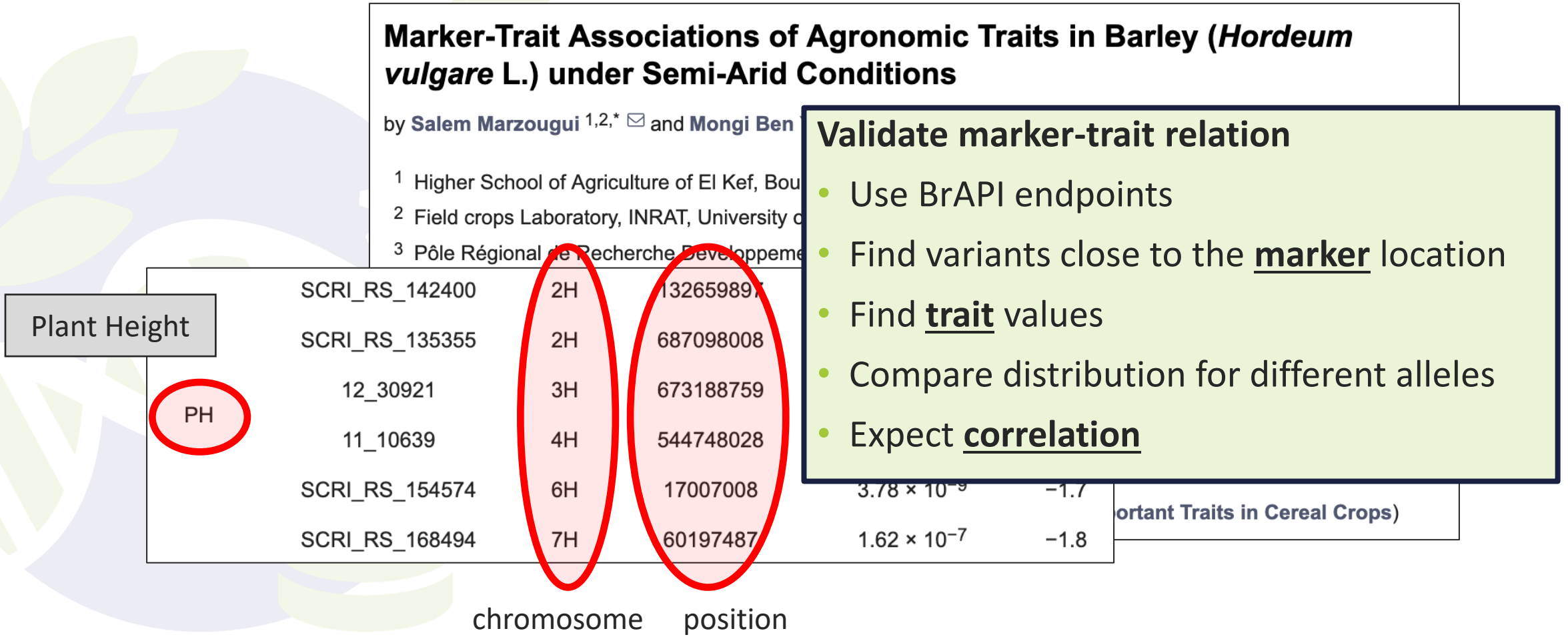
- Merges lists
- Merges (allele)matrices
- Supports prefixing identifiers
- Supports token authentication
- Configurable
- Swagger interface

However

- No support for sorting when merging multiple endpoints
- Assumes correctly implemented BrAPI endpoints

> **Documentation**
> It is highly recommended that every BrAPI server have a documentation page showing the available BrAPI endpoints. Each server has special requirements, additional fields, and unsupported parameters. It is important to have a place to record these small deviations from the specification. If several groups have the same deviation from the spec, the community might decide to add that as part of the next version of BrAPI.

Example: validate marker-trait association from literature

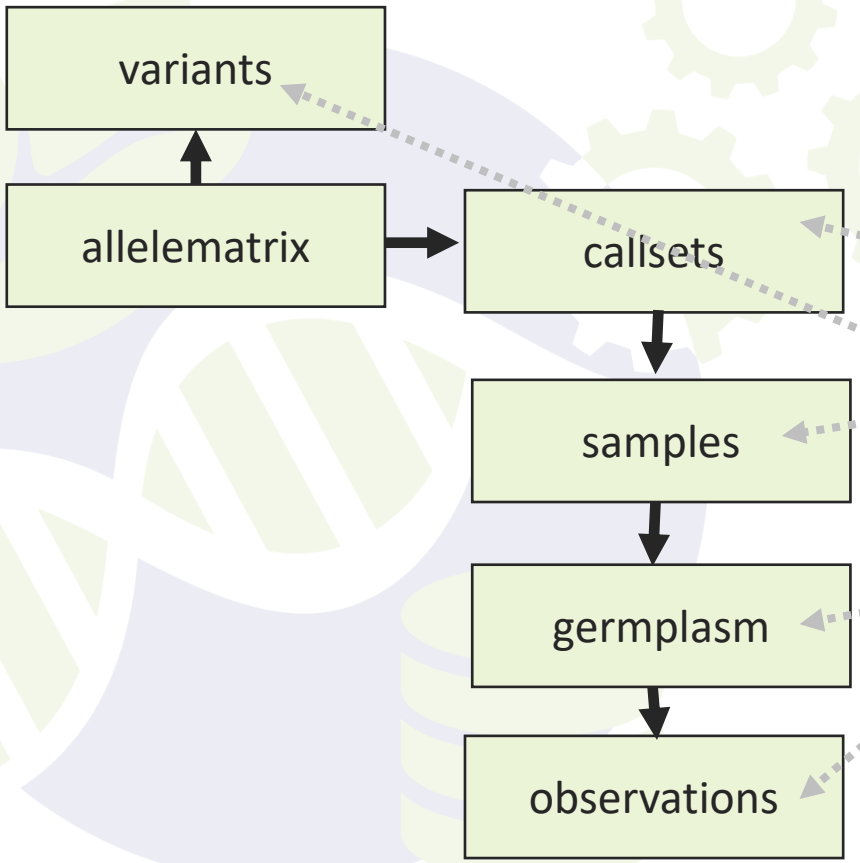


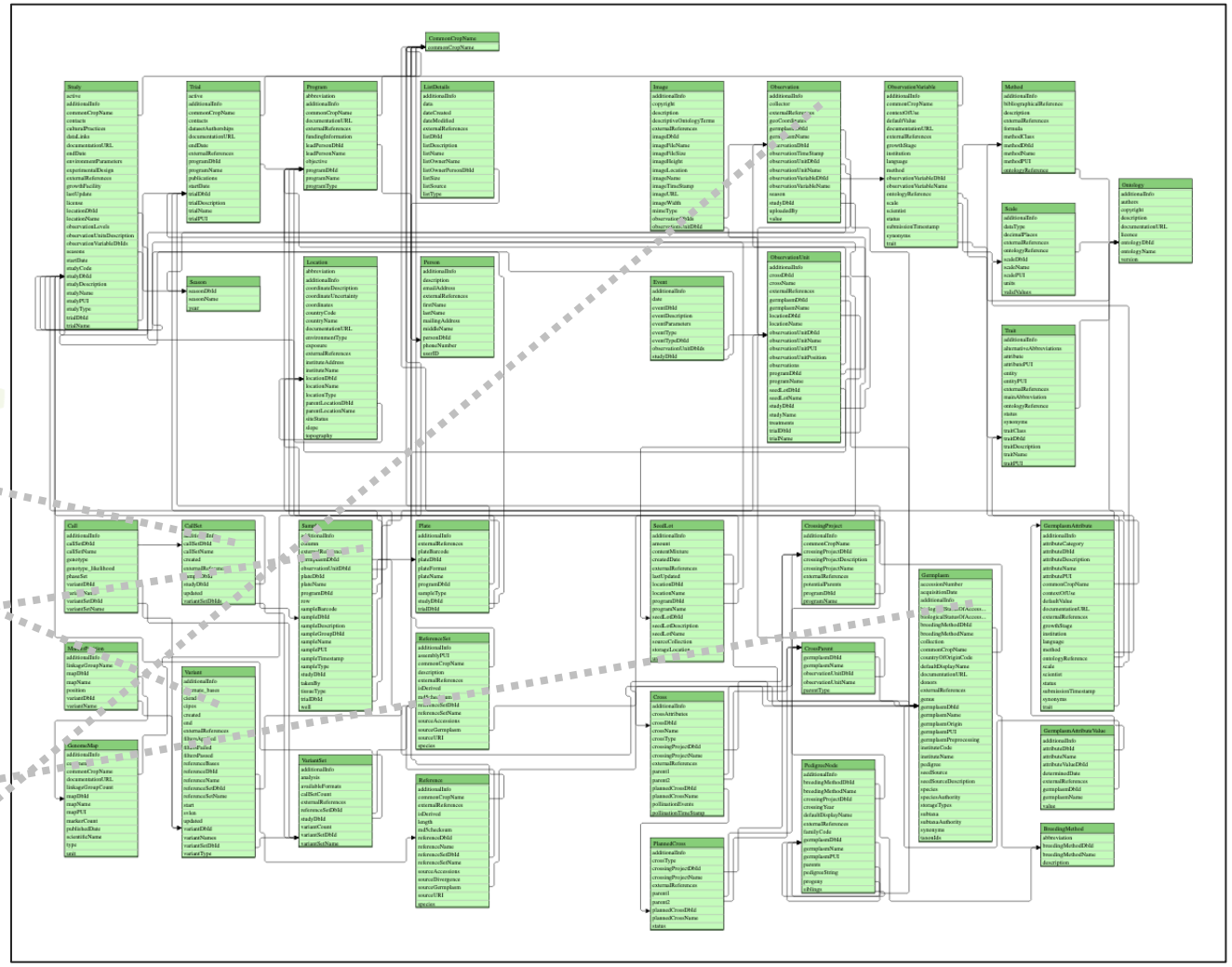
Validate marker-trait relation

- Use BrAPI endpoints
- Find variants close to the **marker** location
- Find **trait** values
- Compare distribution for different alleles
- Expect **correlation**

BrAPI model

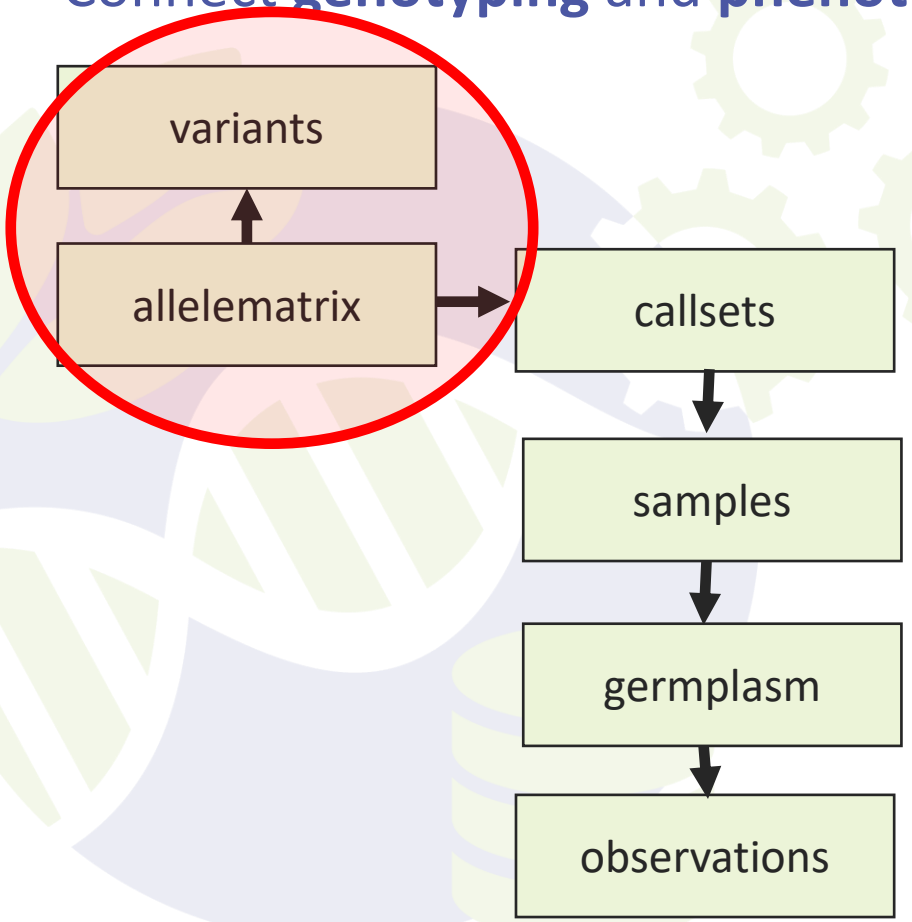
Validate marker-trait relation





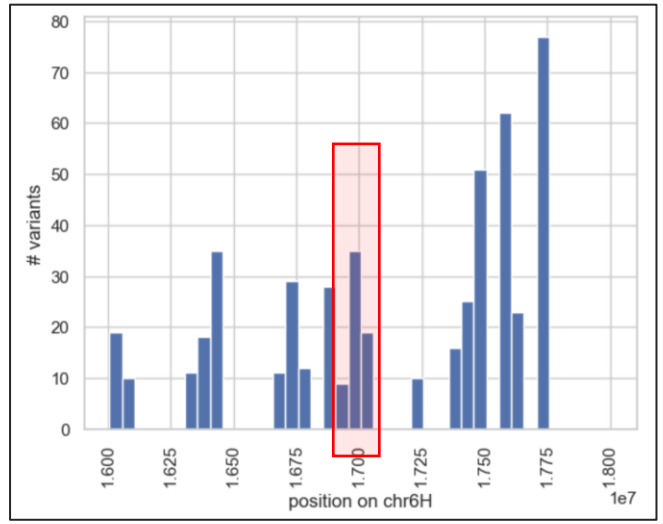
BrAPI model

Connect **genotyping** and **phenotyping**



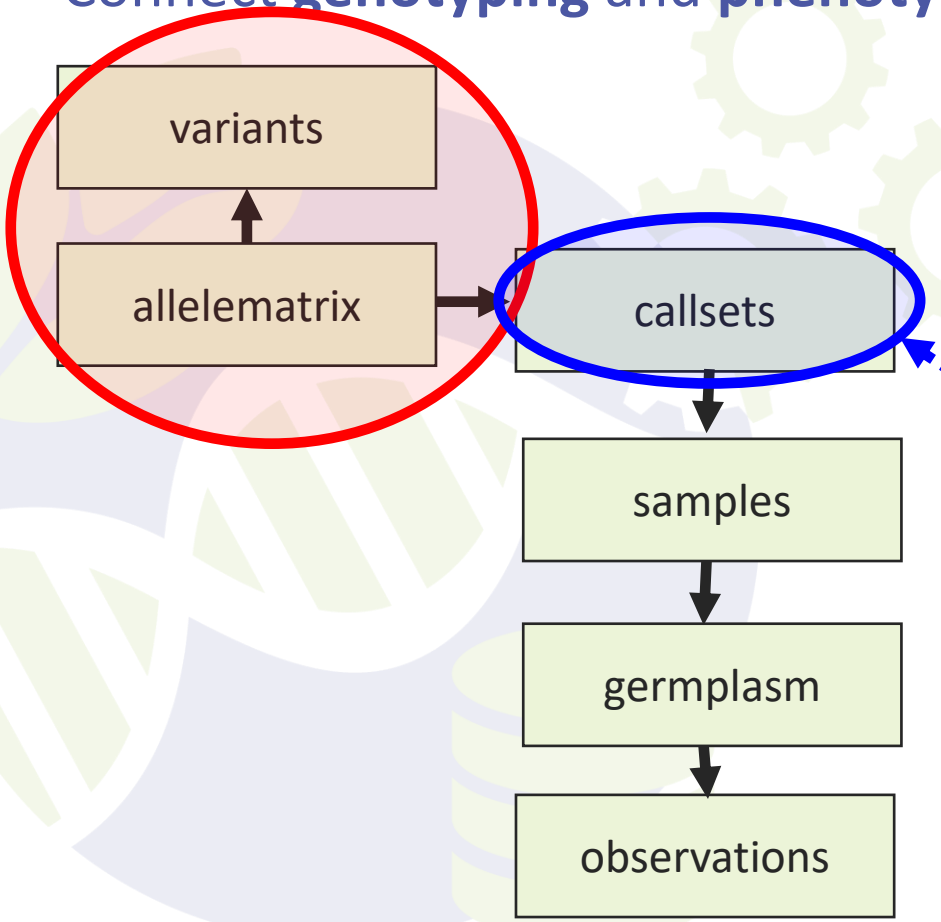
PH	SCRI_RS_142400	2H	132659897	1.89×10^{-6}	−1.89
	SCRI_RS_135355	2H	687098008	6.65×10^{-6}	−1.24
	12_30921	3H	673188759	1.57×10^{-6}	−1.32
	11_10639	4H	544748028	2.05×10^{-6}	1.48
	SCRI_RS_154574	6H	17007008	3.78×10^{-9}	−1.7
	SCRI_RS_168494	7H	60197487	1.62×10^{-7}	−1.8

- use BrAPI endpoints to collect variants on **chr6H**
- select variants around position 17007008
- use BrAPI endpoints to collect genotypic data for selected variants



BrAPI model

Connect **genotyping** and **phenotyping**



PH	SCRI_RS_142400	2H	132659897	1.89×10^{-6}	−1.89
	SCRI_RS_135355	2H	687098008	6.65×10^{-6}	−1.24
	12_30921	3H	673188759	1.57×10^{-6}	−1.32
	11_10639	4H	544748028	2.05×10^{-6}	1.48
	SCRI_RS_154574	6H	17007008	3.78×10^{-9}	−1.7
	SCRI_RS_168494	7H	60197487	1.62×10^{-7}	−1.8

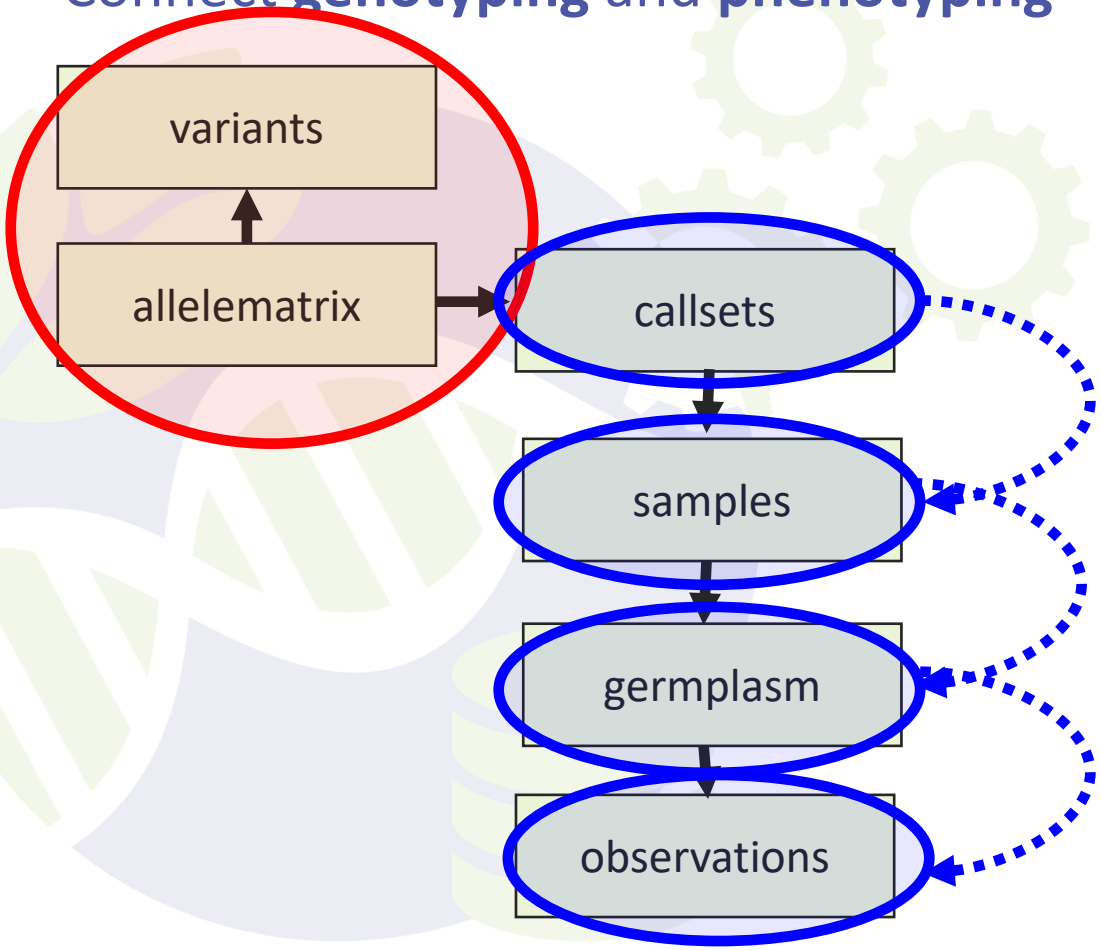
selected variants around 17007008 on chr6H

callset selection

0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0
0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0
1/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0
0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0
0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0

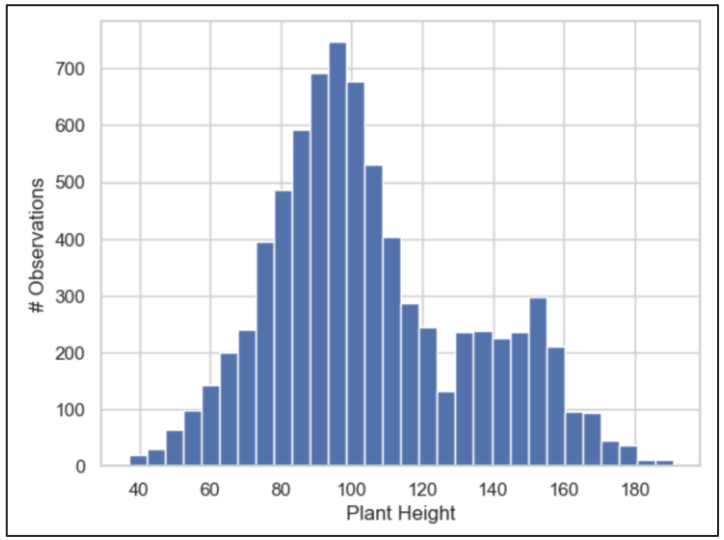
BrAPI model

Connect **genotyping** and **phenotyping**



PH	SCRI_RS_142400	2H	132659897	1.89×10^{-6}	−1.89
	SCRI_RS_135355	2H	687098008	6.65×10^{-6}	−1.24
	12_30921	3H	673188759	1.57×10^{-6}	−1.32
	11_10639	4H	544748028	2.05×10^{-6}	1.48
	SCRI_RS_154574	6H	17007008	3.78×10^{-9}	−1.7
	SCRI_RS_168494	7H	60197487	1.62×10^{-7}	−1.8

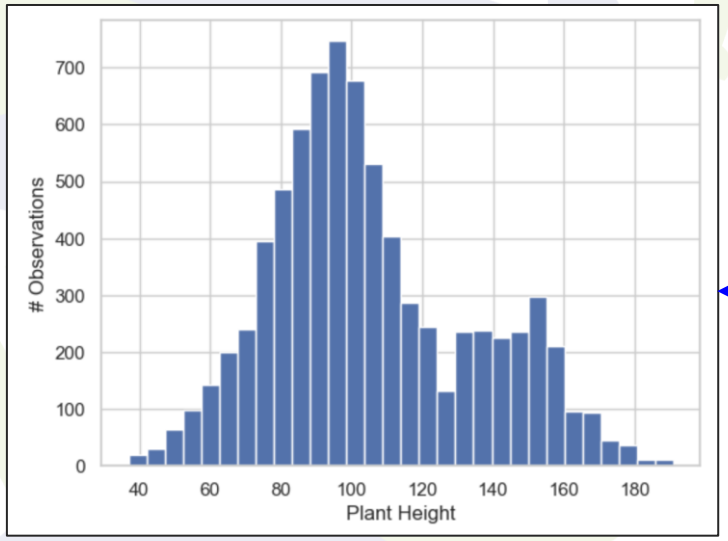
- use BrAPI endpoints to find observations linked to these callsets
- restrict to observations for “Plant Height”
- 7710 values for 3310 unique germplasm



BrAPI model

Connect **genotyping** and **phenotyping**

PH	SCRI_RS_142400	2H	132659897	1.89×10^{-6}	−1.89
	SCRI_RS_135355	2H	687098008	6.65×10^{-6}	−1.24
	12_30921	3H	673188759	1.57×10^{-6}	−1.32
	11_10639	4H	544748028	2.05×10^{-6}	1.48
	SCRI_RS_154574	6H	17007008	3.78×10^{-9}	−1.7
	SCRI_RS_168494	7H	60197487	1.62×10^{-7}	−1.8



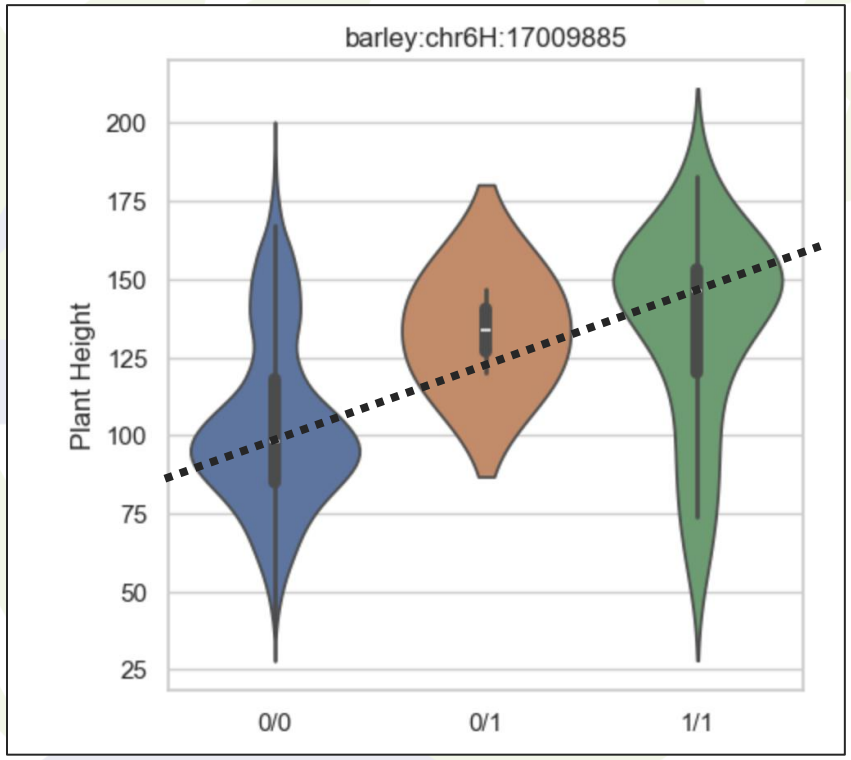
selected variants around 17007008 on chr6H

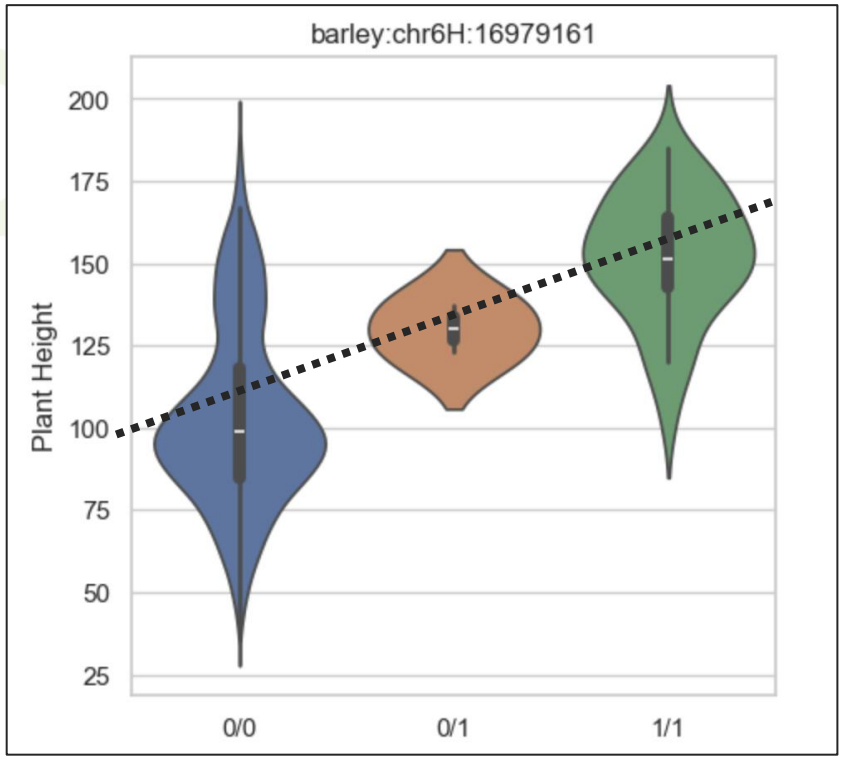
callset selection

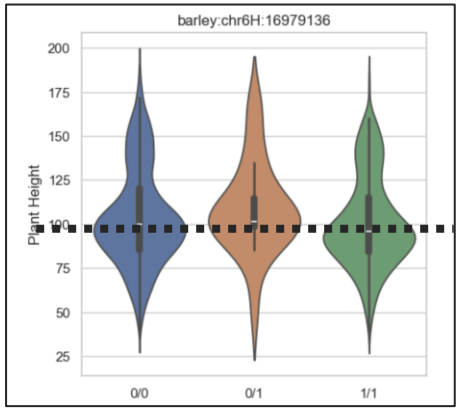
	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/1	0/0
	0/0	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0
	1/1	0/0	0/0	0/1	0/0	0/0	0/0	0/0	0/0
	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/0	0/0
	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0	0/0
	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0
	0/0	0/0	0/0	1/1	0/0	0/0	0/0	0/0	0/0

Example: validate marker-trait association from literature

Phenotypic distribution for different genotypic values

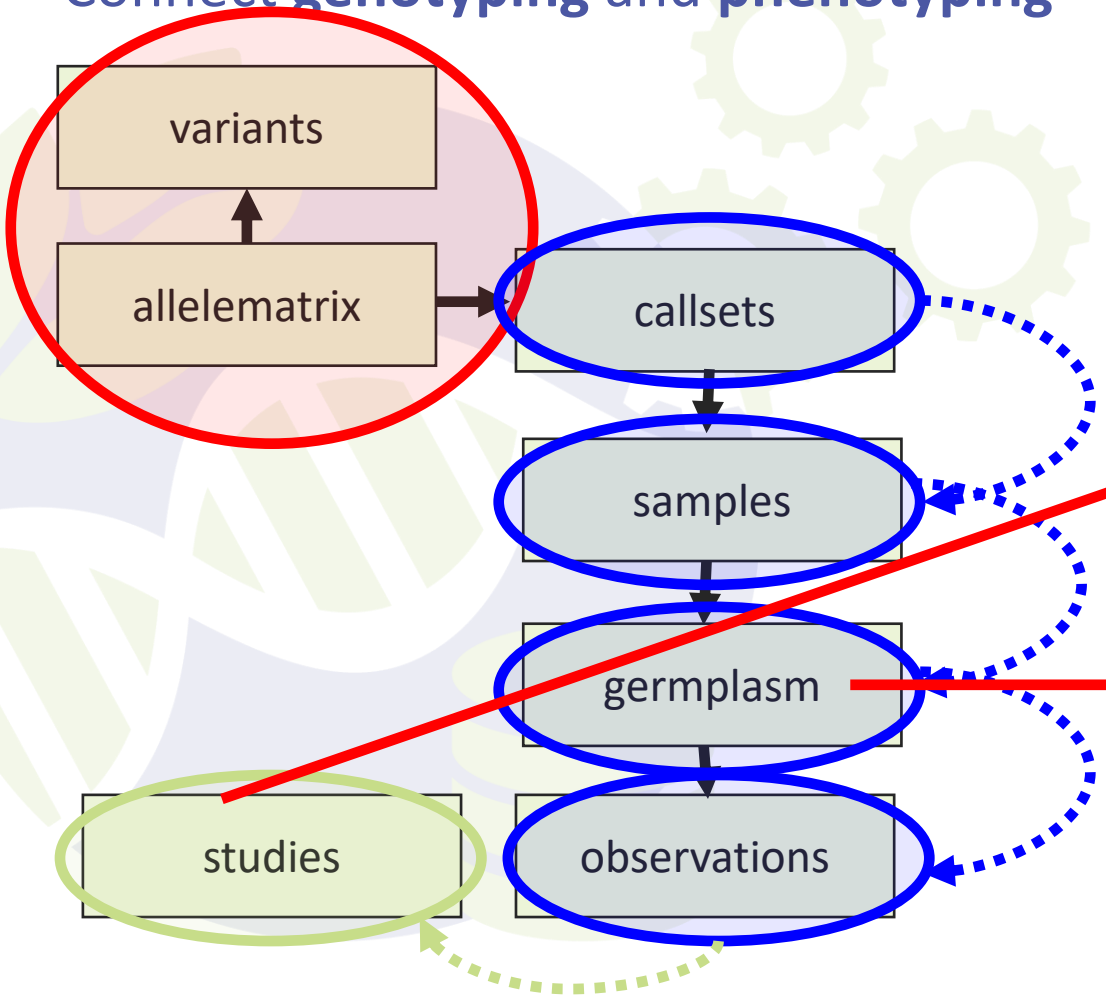


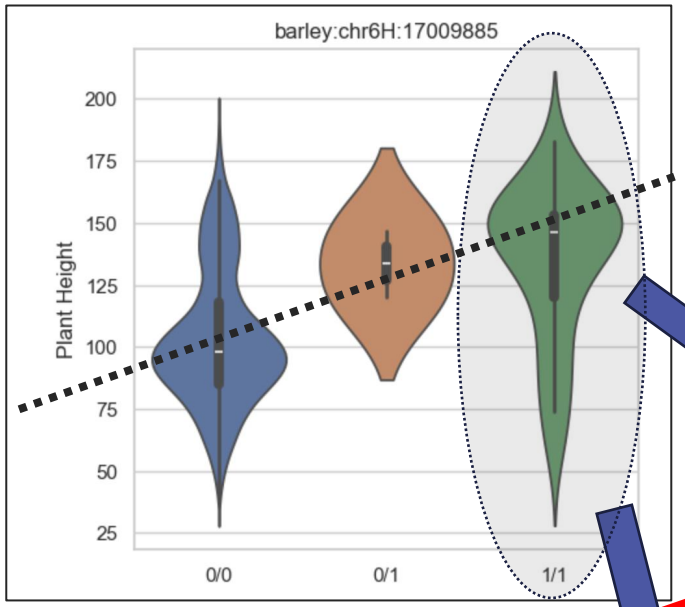


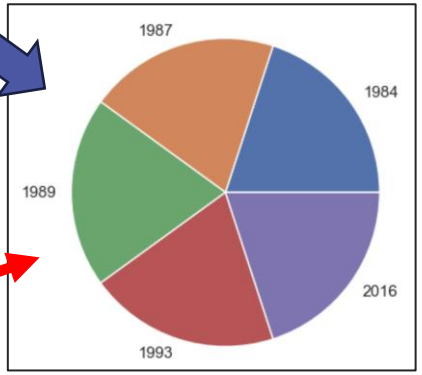


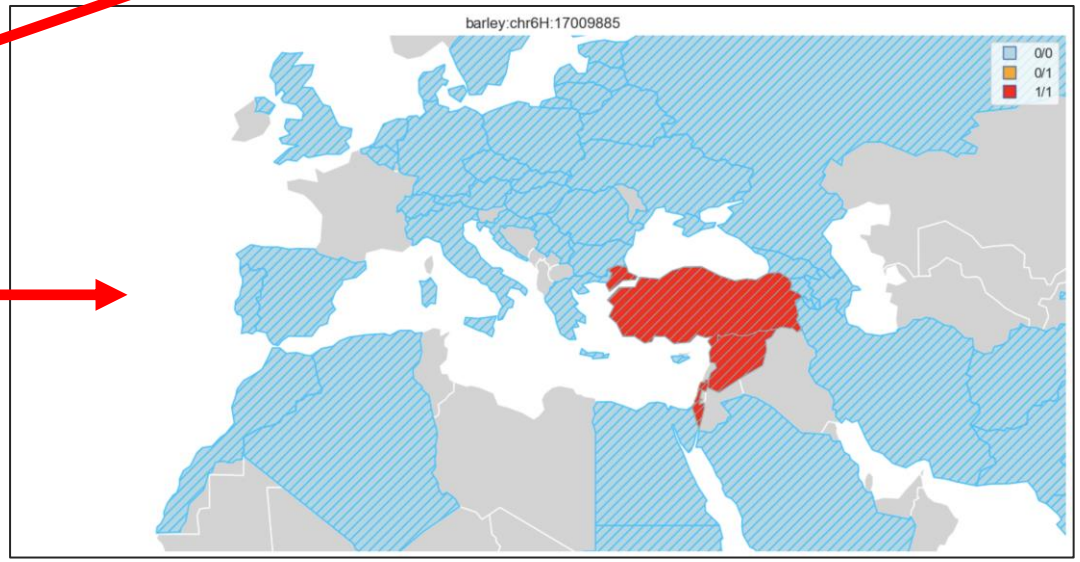
BrAPI model

Connect **genotyping** and **phenotyping**









Conclusion

- **AGENT API**: A programming interface for data
- Adoption of Breeding API - **BrAPI**
- Data Storage within the AGENT Project: usage **brapi-proxy**
- **Example**: validate marker-trait association from literature

Thank you!

Contact Information: matthijs.brouwer@wur.nl

- Matthias Lange
- Cyril Pommier
- Gouripriya Davuluri
- Joseph Ruff
- Patrick König
- Paul Kersey
- Stephan Weise